

ABSTRACT

Title of Thesis:

A NEW APPROACH FOR FINE MAPPING
AND CLONING GENES FROM THE
TERTIARY GENE POOL MEMBERS OF
WHEAT: EXAMPLE OF *Lr57* FROM
AEGILOPS GENICULATA

James Steadham, Master of Science, 2020

Thesis Directed By:

Dr. Nidhi Rawat, Department of Plant Science
and Landscape Architecture

Wheat is a staple food for 35% of the world's population, providing more calories and protein to the world's diet than any other crop. Wheat rust diseases cause yield losses worth billions of dollars annually. Two rust resistance genes, *Lr57* and *Yr40*, conferring leaf rust and stripe rust resistance, respectively, were previously identified in a wild wheat relative, *Aegilops geniculata*. Mapping these genes within a wheat background is typically impossible, due to the absence of genetic recombination between wheat and wild species chromatin. We devised a novel technique to overcome this barrier, developing a mapping population from a cross between a resistant *Ae. geniculata*-wheat introgression line and a susceptible *Ae. geniculata* disomic addition line. Genotyping and phenotyping 162 individuals from this population, we found 11 recombinants within a 6 Mb interval and fine mapped *Lr57*, demonstrating the high-resolution power of this strategy for mapping genes from wild relatives of wheat.

A NEW APPROACH FOR FINE MAPPING AND CLONING GENES FROM THE
TERTIARY GENE POOL MEMBERS OF WHEAT: EXAMPLE OF *LR57* FROM
AEGILOPS GENICULATA

by

James Steadham

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science

2020

Advisory Committee:

Dr. Nidhi Rawat, Chair

Dr. Vijay Tiwari

Dr. James Culver

© Copyright by

James Steadham

2020

Dedication

Dedicated to my parents, my sister, my mother-in-law, my brother-in-law, and most importantly my wife, Morgan.

Acknowledgements

To loosely paraphrase something from my father's dissertation: having written a thesis, I have a profound appreciation for what it must take to both read and edit one. Dr. Vijay Tiwari and Dr. Nidhi Rawat went through an astonishing amount of effort editing my thesis, and at the same time had two other students' theses they were working on as well. I don't know how they do it, but they both have my special thanks and admiration.

This work would not have been possible without the help and support of multiple people. Taylor Schulden's help was invaluable in the last month of this project. High quality rust phenotyping and physical map development would not have been possible without the help of Bhanu Kalia and Shichen Wang, respectively.

Finally thank you to the greenhouse staff, Sydney Wallace and Meghan Fisher-Holbert. And thank you to all the undergrad students who've passed through our lab, and the other members of our lab as well: Alex Mahlandt, Lovepreet Singh, Bhavit Chhabra, and Adam Schoen.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	v
List of Figures	vi
Chapter 1: Introduction	1
Background	1
The Three Gene Pools of Wheat	3
The <i>Phl</i> Locus	4
Rust Disease of Wheat	6
The <i>Aegilops</i> Species and Rust Resistance	7
The <i>Aegilops</i> species	7
Project Outline	10
Chapter 2: Literature Review	13
Rust Infection, The MAMP Response, and Disease Resistance	13
Disease Resistance Genes in Wheat and Cloning	16
<i>Aegilops. geniculata</i> , TA5602 and TA5601	17
<i>Aegilops umbellulata</i>	17
Chapter 3: An Efficient Approach to Precisely Map Genes of Non-Recombining Chromosomes from Distant Gene Pools	19
A Novel Mapping Approach	19
Materials and Methods	25
Results	35
Discussion	35
Chapter 4: Physical Mapping and Identification of an Lr57 Candidate Gene	56
Abstract	56
Introduction	58
Methods	61
Results	64
Discussion	69
Bibliography	77

List of Tables

Chapter 3:

Table 3.1: Plant Materials used

Table 3.2: Leaf rust

Table 3.3: Stripe rust

Table 3.4: List of primers for characterizing TA5602

Table 3.5: List of primers for characterizing TA5601

Table 3.6: Lr57 and Yr40 segregating phenotypes

Table 3.7: Primers for mapping Lr57

Table 3.8: Marker scores as tested on the $F_{2:3}$ population

Chapter 4:

Table 4.1: Transcription differences with TA5602 and WL711 aligned to wheat

Table 4.2: Comparing favorable mapping conditions for TA5602 and WL711

List of Figures

Chapter 1

Figure 1.1: *Aegilops geniculata*

Figure 1.2: An overview of the process of creating 5M^g specific markers

Figure 1.3: Chromosomes 5M and 5M^g of our plant material. The 5M^g cross of TA659 and TA5602

Chapter 2:

Figure 2.1: rust lifecycle

Chapter 3:

Figure 3.1: Dominant marker creation

Figure 3.2: Characterizing the 5M^g introgression size of TA5602

Figure 3.3: Characterizing the 5M^g introgression size of TA5601

Figure 3.4: A unique mapping scheme

Figure 3.5: Cytological analysis of pairing

Figure 3.6: Leaf rust phenotyping

Figure 3.7: Stripe rust phenotyping

Figure 3.8: An example of a SNP based marker

Figure 3.9: Genetic map of *Lr57* candidate region

Chapter 4:

Figure 4.1: A representation of the mapped TA5602 introgression size as represented by the physical and genetic map

Chapter 1: Introduction

Background

Over 10,000 years ago, with the transition of the earth's climate from the Pleistocene (ice age) to the current Holocene era, so too were human societies then able to transition from nomadic to agrarian lifestyles- a shift fundamental for accommodating human population growth to over 7 billion today (Barker & Goucher, 2015). Einkorn wheat (*Triticum monococcum*) and emmer wheat (*Triticum turgidum* subsp. *dicoccum*) were founder crops, two of the first crops domesticated by humans in the Fertile Crescent, in part due to their loss of a brittle rachis that would have allowed free seed dispersal and made harvesting difficult (Faris, 2014; Weiss & Zohary, 2011). In this same region some 8,500 years ago, allohexaploid bread wheat (*Triticum aestivum* L., $2n = 6x = 42$, AABBDD) developed as the result of a hybridization event between *T. turgidum* L., ($2n = 4x = 28$, AABB) and *Aegilops tauschii* Coss. ($2n = 2x = 14$, DD) (Faris, 2014). Due to hexaploid wheat's rare hybridization origin, the origin being confined to the Fertile Crescent, and a subsequently long history of breeding within wheat's immediate gene pool, wheat is particularly influenced by the Founder effect and has a narrow genetic base for necessary future breeding improvements (Faris, 2014; Ladizinsky, 1985; Tiwari et al., 2014).

To stress the importance of breeding programs, according to data from 2017, bread wheat is the most cultivated crop, contributes nearly a fifth of human calorie

intake, and is the greatest protein provider (Consortium (IWGSC) et al., 2018; *FAOSTAT*. 2020, Jun 7. <http://www.fao.org/faostat/en/#data/FBS>; *FAOSTAT*. 2020, Jun 7. <http://www.fao.org/faostat/en/#data/QC>). Most of this consumption is within developing countries that consume 77% of the wheat produced (Enghiad et al., 2017; 2020, Jun 7. <https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery>. These same countries are experiencing increased population growth, increased wheat consumption, and increasing wheat prices (Enghiad et al., 2017; *USDA ERS - Wheat Data*. 2020, June 7. <https://www.ers.usda.gov/data-products/wheat-data/>). Failure to meet yield demands could have devastating consequences. In the modern day Fertile Crescent country Syria, a three year long drought induced by climate change caused the country's "breadbasket" to collapse, contributing to the ongoing Syrian civil war (M. Ali, 2010; Kelley et al., 2015; Trigo et al., 2010). Following the drought, epidemics of stripe rust between 2010 and 2011 caused further major losses of Syrian wheat yield (Amil et al., 2020).

Since 1960, the world's population has become two and a half times larger, and cereal production has become three times greater- largely due to well-funded advances in crop genetics (Pingali, 2012; Wik et al., 2008). The Green Revolution between 1966 and 1985, and the following decades of crop research until 2000, led to wheat yields in developing countries overall increasing 208%, but certain countries, like those in Africa, largely did not reap the same benefits (Pingali, 2012; University & Organization, 2004). In addition to policy choices, a single gene that caused dwarfing, *Reduced height (Rht)*, by suppressing gibberellin activity was crucial in developing wheat that could support higher grain yields (Hedden, 2003). Continued development

of high-quality wheat cultivars with good yields and disease resistance is of vital importance. The world population is predicted to continue growing from today's 7.7 billion to 9.7 billion in 2050, with rates highest in sub-Saharan Africa and then central and southern Asia (United Nations et al., 2019). In that time, demand for cereals could increase anywhere from 26% to 68%, while challenges to mitigate environmental issues such as agricultural greenhouse gas emissions and nutrient pollution are likely to continue (Hunter et al., 2017).

Due to wheat's genetic bottleneck, efficiently developing new cultivars to meet current and future challenges in part becomes a question of how to expand the range of available genetic resources that can be used in wheat breeding programs (Tiwari et al., 2014). Alien introgression of genetic traits from wild wheat relatives is a highly valuable answer to this question, with "wild" referring broadly to the wheat-related species that can be categorized into three different gene pools based on the constitution of their genomes (Jiang et al., 1993).

The Three Gene Pools of Wheat

Wheat landraces and wheat's progenitors *T. turgidum* L. and *Ae. tauschii* constitute wheat's primary gene pool, its closest relatives with one or more shared genomes. Various *Triticeae* species with homoeologous genomes related to the A, B, or D genomes of wheat make up wheat's secondary gene pool. This includes the known parent of *T. turgidum* L., *T. uratu* Tuanian ex Gandilyan ($2n = 2x = 14$, AA); *Ae. speltoides* Tausch ($2n = 2x = 28$, SS), the predicted closest relative of the unknown B-genome donor; various A-genome species: *T. monococcum* L. subsp. *monococcum*,

aegilopoides, *T. timopheevii* Zhuk. ($2n = 4x = 28$, AAGG); and the D genome cluster present in polyploid *Aegilops* species (Qi et al., 2007; Tiwari et al., 2014). Transfer of genes from the primary and secondary gene pool can be done through traditional crosses based on homologous recombination (Qi et al., 2007). The tertiary gene pool members are *Triticeae* species with no A, B, or D genome, and are recalcitrant to recombination with wheat chromosomes.

Many agronomically important traits have been identified in wild wheat relatives. Resistance genes derived from wild wheat species constitute nearly half the catalogued resistance genes (Bansal et al., 2017). But incorporation of homoeologous breeding material, particularly that of the tertiary gene pool, has been underutilized for several reasons. First, breeding a gene from a wild relative of wheat into an elite cultivar carries greater risks of linkage drag if the exact gene of interest has not been characterized. Second, there are the restrictions imposed by the dominant Pairing homoeologous1 (*Ph1*) locus located on 5BL which developed during wheat's polyploidization history (Okamoto, 1957; Riley & Chapman, 1958; Sears & Okamoto, 1958).

The *Ph1* Locus

Ph1 promotes homologous pairing within A, B, and D genome chromosomes and thus prevents non-homologous recombination among homoeologues (A-B, B-D, or A-D). *Ph1* is responsible for causing polyploid wheat to undergo diploid-like chromosome pairing where 21 bivalents are formed during metaphase 1 (Riley & Chapman, 1958). To prevent nondisjunction, at least one chiasma needs to form per bivalent (Zickler & Kleckner, 1999). Absent the *Ph1* locus, there are then induced

chromosomal rearrangements that affect wheat fertility (Sánchez-Morán et al., 2001). Within the *Ph1* locus, there are clusters of cyclin-dependent kinase (Cdk)-like genes, methyl transferase genes, and a duplicated chromosome 3B segment that contains a single copy of the gene *TaZIP4-B2* (Al-Kaff et al., 2008; Griffiths et al., 2006; Martín et al., 2017; Rey et al., 2017). EMS mutants for *TaZIP4-B2* in the wheat line Cadenza crossed with the wild relative *Ae. variabilis* indicated the gene's involvement in promoting/restricting homologous/homeologous crossing over (Rey et al., 2017).

But although it has previously been assumed the *Ph1* locus restricts homoeologues pairing, in a study of wheat-rye hybrids (which only have homoeologues) by Martín, et. al, it was shown that *Ph1* did not affect homoeologous synapsis prevalence (Martín et al., 2014, p. 1). However, it was also shown that *Ph1* promotes early homologous synapsis, which can reduce the likelihood of later homoeologous synapsis (Martín et al., 2017).

Given the constraints of the *Ph1* locus regarding alien introgression via homoeologous recombination, various methods have been used in the past to lessen its effects, such as *ph1* mutants, nulli-5B, and crosses containing *Ae. speltooides* genes discovered to be epistatic to *Ph1* (referred to as *Ph^l* genes) (Aghaee-Sarbarzeh et al., 2002.; Riley et al., 1961). Notably Chen et. al, 1994, transferred *Ph^l* from *Ae. speltooides* to Chinese Spring (CS), and this showed homoeologous pairing in the F₁ generation in a cross between CS(*Ph^l*) stock and wild *Haynaldia villosa* chromosomes (Chen et al., 1994).

Rust Disease of Wheat

There are three wheat rusts: stem/black (*Puccinia graminis* f. Sp. *tritici* (Pgt)), leaf/brown (*P. triticina* (Ptr), and stripe/yellow rust (*P. striiformis* f. Sp *tritici* (Pst)). Each is a type of obligate biotrophic fungi in the Basidiomycete family. Global annual losses to rust is estimated to be between 4.3 to 5 billion USD (P. Pardey, University of Minnesota, unpublished) (Figueroa et al., 2018).

Currently some 88% of wheat is susceptible to stripe rust and as the range of stripe-rust susceptible environments has increased to include warmer areas, there has been an estimated 90% chance that at least 4.7 million tons of wheat yield will be lost per year (Beddow et al., 2015). Leaf rust is the most common and widely distributed rust (Huerta-Espino et al., 2011). Although leaf rust may affect yields less than stem or stripe rust, due to its spread it may still cause greater annual losses (Huerta-Espino et al., 2011). Both stripe and leaf rust cause yield decreases through lower kernel weight.

Discovering disease-resistance genes for integration into elite cultivars is an important rust management strategy that does not carry the same environmental, cost, and resistance risks of using fungicides against rust (Ellis et al., 2014; Oliver, 2014). Furthermore, multiple genes should be incorporated in cultivars for disease resistance at once, to avoid single *R* gene breakdown and subsequent loss of disease defenses.

The *Aegilops* Species and Rust Resistance

The *Aegilops* Species

Overall the *Aegilops* taxa contains 11 diploid and 12 polyploid species (Kilian et al., 2011). *Ae. geniculata* (Figure 1.1) is an allotetraploid ($2n = 4x = 28$, $U^G U^G M^G M^G$) and one member of the tertiary gene pool that contains many genes that would be useful for wheat improvement, including greater protein, iron, and zinc content, drought and heat resistance, and other traits including disease resistance (Bs et al., 1985; B. Friebe et al., 1996; Rawat et al., 2008; Schneider et al., 2008; Zaharieva et al., 2001). There are reports of high *Ae. geniculata* resistance to stem and leaf rust going back at least as far as 1918 (Valkoun et al., 1985). It was shown in the early 1990's to 2000 that *Aegilops* chromosomes U and M are both sources of rust resistance, and in particular $5M^g$ could be a source of leaf and stripe rust resistance (Dhaliwal, unpublished) (Aghaee-Sarbarzeh et al., 2002; Dhaliwal et al., 1993; Khem Singh Gill et al., 1995; Harjit & Dhaliwal, 2000). By 2018, 75 leaf rust resistance genes had been identified and a similar number of stripe rust genes as well, with ~20% of the genes being derived from *Aegilops* species (Kishii, 2019; Ponce-Molina et al., 2018).



Figure 1.1 *Aegilops geniculata* Roth. There are two subspecies: subsp. *geniculata* and subsp. *gibberosa* (Zhuk.) Hammer. This thesis refers to the former subspecies that originated in the eastern Mediterranean, while the other is localized to Spain and North Africa (Arrigo et al., 2010; Ohta, 2017).

There have been several reports of hexaploid wheat hybridizing with *Ae. geniculata* despite the *Ph1* locus, first by Requier in 1825 by Avignon in southern France and then by Fabre of Agde, France in 1838 (Dondlinger, 1919; Loureiro et al., 2006). Between 2003 and 2004 in Madrid Spain, simulating field conditions, Loureiro, et. al showed more conclusively a hybridization rate of between .24% and .39% between hexaploid wheat and *Ae. geniculata* (Loureiro et al., 2007). Similarly, Koo et al., showed evidence of a 3.4% (n=114) homoeologous pairing rate between 5M^g and 5D in F₁ hybrids of *Ae. geniculata* and wheat and saw high chiasmatic associations between M genomes and wheat, suggesting the potential for some homoeologous promoting gene(s) on 5M^g (Koo et al., 2017).

Given the low rate of normal hybridization between *Ae. geniculata* and wheat, CS *Ph1* was an important tool in a study by Aghaee-Sarbarzeh et al., 2002 to transfer a small segment *Ae. geniculata* 5M^g chromatin into a hexaploid wheat background, along

with the leaf and stripe rust resistance the segment of 5M^g chromatin provided (Aghaee-Sarbarzeh et al., 2002). Specifically, this involved first creating a disomic substitution line DS5M^g(5D) from a cross between the rust-resistant *Ae. geniculata* accession TA10437 and the highly rust-susceptible wheat variety WL711. Crossing DS5M^g(5D) with CS *Ph1* and then crossing the F₁ with WL711 led to homoeologous recombination and the creation of wheat-*Ae. geniculata* introgression lines.

***Lr57* and *Yr40*: The First Characterization**

Wheat-*Ae. geniculata* introgression lines had been originally developed using CS *Ph1* stock and a disomic substitution line DS 5M^g(5D) that was backcrossed to WL711 by Aghaee-Sarbarzeh et al., (2002). The resulting BC₁F₁ plants were then crossed and backcrossed to the highly rust susceptible soft white spring wheat cultivar WL711 by Kuraparthi et al., (2007). These resulting stable introgression lines included TA5599 (BC₂F₅), TA5601 (BC₂F₅), and TA5602 (BC₃F₆) that, through genomic *in situ* hybridization (GISH) against *Ae. comosa* and Restriction Fragment Length Polymorphism (RFLP) probes for orthologous alleles in wheat 5A/5B/5D, showed 5M^g replacing either the majority of wheat chromosome 5D (TA5599 [T5MS·5ML-5DL]), approximately 25% of 5D (TA5601 [T5DL·5DS-5M^gS(0.75)]), to under 5% of wheat 5DS (TA5602 [T5DL·5DS-5M^gS(0.95)]). TA5602 was further described as a cryptic introgression, due to the 5M^g introgression size being cytologically “invisible” while still containing *Lr57* and *Yr40* (Kuraparthi, 2007).

Project Outline



Figure 1.2: A generalized overview of the process of creating 5M^g specific markers, creating a mapping population, and identifying a candidate region for our trait of interest.

This thesis project, in broad terms, had the following goals:

- Characterizing the introgression size of TA5602 and TA5601
- Determining if *Lr57* and *Yr40* are one gene or separate genes.
- Fine mapping and identification of *Lr57* candidate genes

Through developing a feasible means of cloning genes from tertiary gene pool species, we aimed to take an important step toward expanding the available genetic resources available to wheat breeding programs.

This project can be generally divided into stages of marker development, testing the markers on a phenotyped mapping population to find a candidate region, and validation of candidate genes (Figure 1.2). In and of itself, this will seem familiar compared to traditional approaches of positional cloning. The real advantage of this project's approach is that it shows how a unique mapping population developed from an introgression line (TA5602) crossed with a disomic addition line (TA7659) can show specific recombination limited only to a small region of interest containing candidate genes, due to the *Ph1* locus. This cuts down candidate regions of interest from an entire chromosome level to simply the size of the alien chromatin translocated into a wheat background (figure 1.3). Because both TA7659 and TA5601 are almost entirely leaf and stripe rust susceptible wheat backgrounds (Chinese Spring and WL711, respectively) in regard to our working rust isolates, we can be sure that any rust resistant phenotypes observed in the mapping population will be due to 5M^g and, in particular, *Lr57* and *Yr40*.

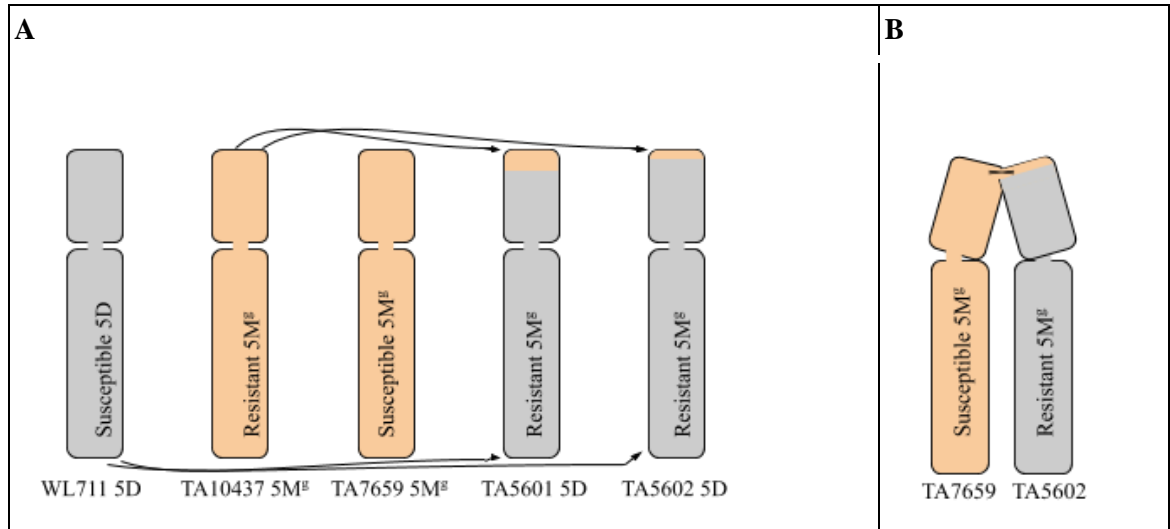


Figure 1.3: Represented in panel **A** are, from left to right: wheat chromosome 5D, leaf and stripe rust resistant 5M^g chromosome of TA10437, leaf and stripe rust susceptible 5M^g chromosome in the disomic addition line TA7659 [DS 5M^g (5D)], and the two introgression lines TA5601 and TA5602 that include a 5M^g translocation segment from TA10437 replacing the distal end of WL711 5D. Panel **B** is a representation of a cross between TA7659 and TA5601, where only the *Ae. geniculata* 5M^g segments with shared homology align and recombine, due to the *Ph1* locus, forming a trivalent. Any rust resistance phenotypes within the progeny of this cross are directly tied to TA5601. *Orange represents geniculata 5M^g chromatin while grey represents wheat 5D chromatin.*

Furthermore, we showed how we can compensate for the lack of genomic resources, such as reference sequences, in wild relatives of wheat, even in the tertiary gene pool, through comparative genomics to better developed sources, such as the wheat reference sequence. This allowed us to easily generate 5M^g specific markers that were crucial in mapping *Lr57* and *Yr40*. The same approach was used to develop markers for both QPCR and creating VIGS constructs to test candidate genes.

Chapter 2: Literature Review

Rust Infection, The MAMP Response, and Disease Resistance

Pucciniales are a monophyletic group and one of the most speciose groups of plant pathogens with approximately 8000 species described (Aime et al., 2017). The cereal rusts all follow macrocyclic life cycles consisting of five different stages and are heteroecious, requiring two unrelated hosts to complete the sexual life cycle and develop diversity quickly (figure 2.1). However, the alternative hosts for leaf rust are, with little exception, not crucial factors in epidemic spread of rust (J. Zhao et al., 2016). Stripe rust has been thought to spread clonally in Australia and Europe and epidemics seemed to be largely due to more divergent lines (S. Ali et al., 2017; Hovmøller & Justesen, 2007; Schwessinger et al., 2020; Steele et al., 2001). Still in a study comparing stripe rust isolates from 1960 to 2005 in the United States, genetically distinct, newly-appearing isolates of stripe rust appeared to rapidly become predominant in eastern states (Markell & Milus, 2008). The actual rust infection on wheat begins with either an asexual aeciospore or urediniospore, with the aeciospore infection marking the completion of the sexual lifecycle while the urediniospore can repeatedly infect more wheat as part of an asexual lifecycle. The urediniospore develops a germ tube, then appressorium, and then an infection peg that enters wheat cells, eventually forming the haustorium (Bakkeren & Szabo, 2019).

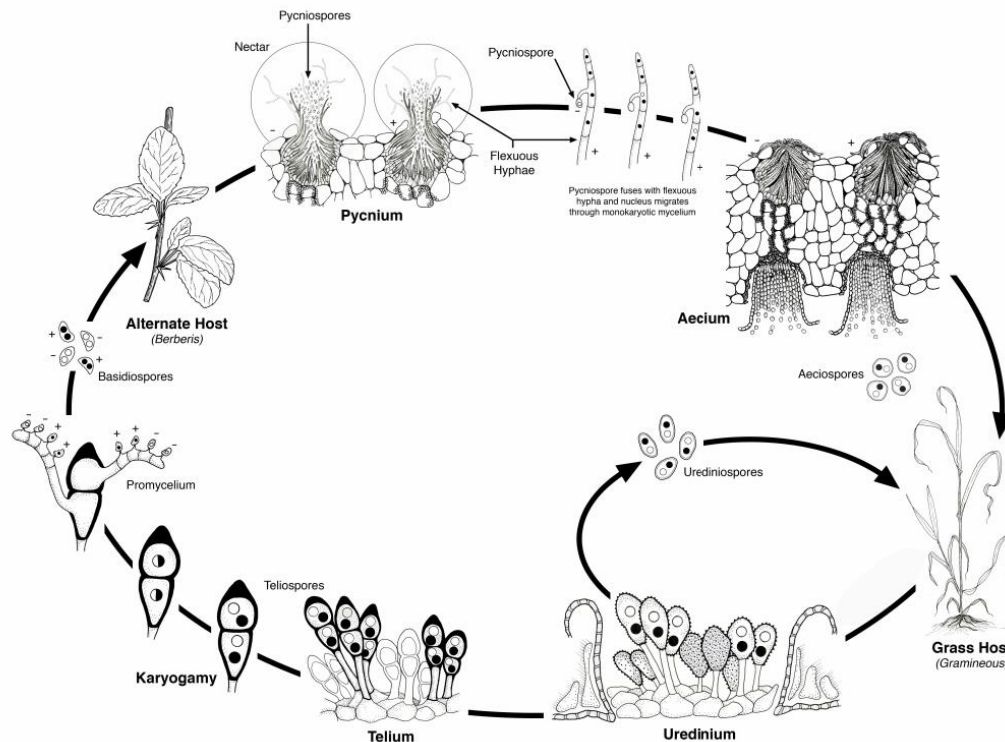


Figure 2.1 An illustration that shows the life cycle of *Puccinia graminis* (Kolmer, 2013). Leaf, stripe, and stem rust each have life cycles that are both heteroecious and macrocyclic.

Drawing by Jacolyn A.Morrison, USDA-ARS Cereal Disease Laboratory, St Paul, MN, USA

Genetic resistance to rust can be broadly divided into seedling and adult-plant resistance. Seedling resistance persists into adult plants and generally shows a hypersensitive response. So far it has been observed that most genes that confer seedling resistance against wheat rust are nucleotide binding and C-terminal leucine-rich repeat (NLR-type) proteins (Schwessinger, 2017). NLR genes can be divided into two different types, based on the N-terminal domain. There are Toll-interleukin-like

receptor (TIR) domains and coiled-coil (CC) domains, the latter type being the only kind known to be present in monocots (Tarr & Alexander, 2009)

The main defense systems of plants are microbe-associated molecular patterns (MAMPs or) or damage-associated molecular patterns (DAMPs) that rely on pattern-recognition receptors (PRRs) in the form of plasma membrane receptor kinases or receptor proteins that can recognize conserved pathogen MAMPs like chitin (Zipfel, 2014). This recognition can lead to downstream defense responses like increased reactive oxygen species production, in what is referred to as PAMP-triggered immunity (PAMPs referring to the MAMPs of pathogens), or PTI. The second defense type of defense response acts mostly within the cell and with NLR genes, where pathogen-produced effector proteins that suppress PTI are in turn suppressed by the plant in what is called effector-triggered immunity, or ETI, and is often associated with a hypersensitive response (Jones & Dangl, 2006). In other words, most of the seedling resistance described in wheat against rust has been a form of ETI. Often stepwise, asexual evolution of rust can then quickly cause virulence against a single NLR gene that had adapted to a specific effector in gene-for-gene fashion (Flor, 1971; Schwessinger, 2017). Additionally, rust genomes are highly complicated and difficult to study, being obligate biotrophs. Rust genomes are large, with stripe rust genomes predicted to be between 50 to 100 Mb, and containing over 1000 potential effectors

(Schwessinger, 2017). As a whole, the average Pucciniales genome size is estimated to be 305.5 Mb (Tavares et al., 2014).

Disease Resistance Genes in Wheat and Cloning

In 1992, *Hm1* from *Zea mays* was the first *R* gene cloned (Johal & Briggs, 1992; Kourelis & Hoorn, 2018). Three years later, the first rust resistance gene was cloned from flax, *L6*, and then in 2003 the first leaf rust resistance gene in wheat was cloned, *Lr21* (Huang et al., 2003). In a survey of 314 plant *R* genes cloned between 1992 and 2018, most of the *R* genes were surface or intracellular receptors and could be divided into nine different categories, but 61% were NLRs (Kourelis & Hoorn, 2018).

In one method of cloning *R* genes, homology-based cloning, where new *R* genes are identified based on homology, four different wheat *R* genes have been identified, all NLR-types (Hurni et al., 2013; Jordan et al., 2011; Mago et al., 2015; M. Wang & Chen, 2017; C. Z. Zhao et al., 2016). In a study analyzing 70 identified *Lr* genes, it was found that NLR genes tended to be positioned close to QTLs for race-specific resistance expressed at the seedling stage (Peng & Yang, 2017). But so far, only three seedling stage leaf rust genes have been cloned (each of the NLR type) and three adult-stage resistant *Lr* genes have been cloned (Prasad et al., 2020).

Aegilops geniculata

Since the original development and characterization of *Ae. geniculata*-wheat introgression lines, multiple studies have made advances in our understanding of both *Ae. geniculata* and the 5M^g chromosome specifically. First, SNPs have been discovered to help track 5M^g introgressions in wheat backgrounds. The leaf and stripe rust susceptible 5MS chromosome derived from TA2899 was flow sorted and sequenced to 45X coverage and used to discover SNPs between 5M^g and wheat group-5 chromosomes (B. R. Friebe et al., 1999; Tiwari et al., 2014). Then a resistant 5M^g chromosome from the substitution line TA6675 (DS5M^g(5D)) was similarly sequenced with paired-end reads and shown to be highly similar to wheat 5D in terms of both synteny and collinearity (Tiwari et al., 2015). It has also now been shown that over 75% of *Aegilops* spp. accessions are resistant to the stem rust races TTKSK, TRTTF, and TTTTF (Olivera et al., 2018). The stem rust gene derived from *Ae. geniculata* known is *Sr53* that shows resistance against the UG99 race TTKSK (W. Liu et al., 2011).

Aegilops umbellulata

The parental species of *Aegilops geniculata* ($2n = 4x$, MMUU) are *Ae. comosa* Sm. in Sibth & Sm. ($2n = 2x = 14$, MM) and *Ae. umbellulata* Zhuk. ($2n = 2x = 14$, UU) (B. R. Friebe et al., 1999). Previously, rust and stripe resistance genes were transferred from the *Ae. umbellulata* accession PAU 3732 into a wheat background to generate the introgression line IL 393-4 (Chhuneja et al., 2008). IL 393-4 was used to create a recombinant inbred line population of 234 lines that appeared to show the *Ae.*

umbellulata rust resistance genes (*Lr76* and *Yr40*) co-segregate (Bansal et al., 2017). It appeared that *Lr76* and *Yr40* in IL 393-4 could be placed with a 9.47 Mb region of 5DS telomeric region and showed complete linkage (Bansal et al., 2020). *Lr76* and *Yr40* also displayed an NLR-like hypersensitive response. Furthermore, a cross between IL 393-4 and TA5602 seemed to show repulsion linkage between *Lr76* and *Lr57* (Bansal et al., 2017). Thus, if *Lr76* and *Lr57* are alleles of each other, cloning either one could simultaneously clone the other.

Chapter 3: An Efficient Approach to Precisely Map Genes of Non-Recombining Chromosomes from Distant Gene Pools

A Novel Mapping Approach

Wild relatives of wheat (“wild” being considered in terms of their homoeologous chromosomes in respect to a domesticated crop species of interest) are rich sources of genetic variability from iron and zinc content to abiotic stress and disease resistance (Neelam et al., 2011; S. Wang et al., 2011). Kuraparthi et al. (2007) characterized a wild wheat relative *Aegilops geniculata* (2n=4x; UUMM) accession TA10437 that shows leaf and stripe rust resistance, due to the genes *Lr57* and *Yr40*. Both of these genes were transferred into a wheat background by crossing CS(*Ph¹*) and TA10437 and then repeatedly backcrossing to the rust-susceptible cultivar WL711 to create the 5M^g introgression lines TA5602 and TA5601 (BC2F5 and BC3F6 for TA5601 and TA5602, respectively) (Aghaee-Sarbarzeh et al., 2002; Chen et al., 1994; Kuraparthi et al., 2007). When the 5M^g introgression lines TA5602 and TA5601 were first characterized, genomic in situ hybridization tests against *Aegilops comosa* (M) seemed to show TA5602 had such a small 5M^g segment that the introgression was referred to as cytologically invisible or cryptic (Kuraparthi et al., 2007). Restriction fragment length polymorphism (RFLP) probes were used to further map the TA5602 introgression relative to CS nullitetrasonic lines for group-5 chromosomes, showing the introgression was located on 5DS and constituted potentially less than 3.5% of 5DS (Kuraparthi et al., 2007). Compared to the other developed introgression lines derived

from the TA10437 and CS (*Ph*¹) cross (including TA5601, which has an alien introgression replacing ~25% of 5DS), TA5602 was of particular interest due to its small introgression size. The smaller the introgression size conferring disease resistance, the less linkage drag of undesirable genes due to that introgression as well.

Cereal crops, including wheat and rice, are known to have a high level of synteny and collinearity (Ahn et al., 1993; Devos & Gale, 2000; Keller & Feuillet, 2000). Of the Poaceae family, rice was of interest in early comparative genomics studies due to its relatively small genome size of 400 Mb and early available reference sequence. Using 16,000 physically bin-mapped expressed sequence tag (EST) to wheat deletion lines, Kuraparthi et al., (2009) showed the 5M⁸ segment introgressed into 5DS of TA5602 to be less than 3.3 centimorgans in length (Kuraparthi et al., 2009; L. L. Qi et al., 2004). It was also suggested that there could be a high degree of collinearity between particularly 12L of rice and the *Lr57* / *Yr40* region (Kuraparthi et al., 2009).

In wheat, recombination primarily occurs in the distal ends of chromosomes, increasing with distance from the centromere toward the telomeres (K. S. Gill et al., 1993; Kulvinder S. Gill et al., 1996; Lukaszewski & Curtis, 1993). The *Ph1* locus promotes homologous pairing and is responsible for polyploid wheat undergoing diploid-like chromosome pairing where 21 bivalents are formed between homologs during metaphase 1 (Martín et al., 2014, 2017; Riley & Chapman, 1958a). Gene mapping is based on recombination and interpreting recombination frequency as a reflection of the distance between genes and markers, with closer genes and markers being more likely to be linked and inherited together. This traditional approach for mapping is complicated when mapping genes within alien introgressions, because the

Ph1 locus prevents homoeologue pairing and thus homoeologous recombination. Simply removing the *Ph1* locus from wheat to allow homoeologous recombination within the alien introgression would lead to chromosomal rearrangements that affect the stability and fertility of wheat (Sánchez-Morán et al., 2001).

Induced homoeologous recombination between wheat and alien chromosomes has been widely used in wheat germplasm programs. A recessive mutant allele of the *Ph1* locus, *ph1b*, reduces the homoeologous recombination barrier and can be used to induce recombination between homeologs. However, elimination of *Ph1* activity requires several rounds of crosses and selfing to create appropriate genetic backgrounds and then to clean up unwanted recombination between wheat chromosomes. These limitations prolong the generation time of gene transfers in wheat from alien germplasm. Nonetheless, this technique has been used successfully to enhance recombination. For example, suppression of *Ph1* was used to introduce stripe rust resistance gene *yr8* into wheat from *Ae. comosa* (Riley et al., 1968). It has been suggested that positive chiasma interference produced by an initial round of recombination would interfere with subsequent recombination events and reduce one's ability to create subsequent smaller translocations (Lukaszewski, 1995). Yet there is an apparent genetic limit of the ability of chromosomes to pair even in the absence of a functional *Ph1* allele. For example, barley and wheat chromosomes pair minimally in a *ph1b* background (Rey et al., 2015). Moreover, success of such events is further reduced if the alien chromosome with gene(s) of interest is evolutionarily rearranged and from that cannot pair with its wheat counterpart.

Another problem in mapping chromosomes of wild wheat relatives is the lack of genomic resources available compared to wheat. This can be solved primarily by developing the necessary resources or comparative genomics. With the advances in sequencing technologies and flow-sorting techniques, this limitation has begun to be addressed by wheat scientists. Tiwari et al. (2015) flow sorted and sequenced chromosome 5M^g with 50 x coverage using paired end reads from a wheat/*Ae. geniculata* disomic substitution line [DS 5M^g (5D)]. Single-gene fluorescence *in situ* hybridization with this assembly showed no major chromosomal rearrangements between 5M^g and wheat 5D (Tiwari et al., 2015). With this information, the well-annotated wheat 5D sequence can be used to reflect the 5M^g sequence in mapping studies, both to facilitate 5M^g-specific marker development and in identifying genes of interest.

The first approach to use wild and related germplasm for wheat improvement dates to the early 19th century, with wheat and rye hybrids (Wilson, 1873). Since then, thousands of wheat germplasm resources have been generated in the form of alien introgression lines, including addition, substitution, and translocation lines (Forsström & Merker, 2001; Friebe, Badaeva, et al., 1996; Friebe, Jiang, et al., 1996; Gill et al., 2006; Schneider et al., 2008). This large trove of germplasm provides a wonderful existing resource that could be used effectively in wheat breeding programs. However, since alien chromosomes do not pair with wheat chromosomes, they are inherited as a big linkage block, and bring many undesirable traits, making them unsuitable for elite cultivars. In this work we aimed to develop an approach to precisely map alien introgressions and genes from wheat's distant gene pool. Initial screenings of

introgression germplasm have shown existing variation among the same chromosome arm of the same species. The wheat-*Ae. geniculata* disomic substitution line TA6675 (donor *Ae. geniculata* accession TA10437) is resistant for leaf rust and stripe rust resistance genes (Kuraparthi et al., 2007), whereas the disomic addition line TA7659 is highly susceptible for the same rust isolates. Decades of work on TA6675 produced the translocation lines TA5601 and TA5602, showing a small 5M^s fragment size of ~25% and 3.5% on the short arm of chromosome 5D of wheat respectively. Both these translocation lines show presence of the leaf and stripe rust resistance genes *Lr57* and *Yr40*.

Because the 5M^s tip of TA5602 and TA5601 will not pair with wheat 5DS, we normally would not be able to identify or precisely map the gene with linked marker information. This also limits our ability to positionally clone agronomically important genes from the secondary and tertiary gene pools of wheat. But by selecting resistant and susceptible introgression lines of the same chromosome, homologous pairing will be induced, and the subsequent recombination events can be used to precisely map targeted traits.

To show the effectiveness of this approach, we used TA5602, which is a source of both the *Lr57* and *Yr40* genes and has less than 3.5% of the translocation from chromosome 5M^s, and crossed it with a leaf and stripe rust susceptible disomic addition line that contains a full 5M^s chromosome lacking *Lr57* and *Yr40* in a wheat background. Using this strategy, a wild type *Ph1* locus will help ensure both the fertility of progeny and that pairing, and recombination, only occurs between segments of homologous alien chromatin and segments of homologous wheat chromatin, allowing

for the generation of an otherwise seemingly typical F₂ mapping population. By combining this approach, with availability of genomic resources that we developed for chromosome 5M^g, we performed high-resolution mapping of the 5M^g segment containing agronomically important genes in wheat. This approach can be applied in any lab and will open new avenues to perform high-resolution mapping of genes from alien chromosomes in wheat.

Genetic Mapping *Lr57* Conferring Leaf Rust Resistance

Positional cloning can broadly be divided into a fine mapping step and a candidate gene validation step. Traditional fine mapping by creating a mapping population is, however, complicated in undomesticated species that have been less well characterized, may lack genetic resources, and have no physical map. Here we describe the process of mapping the *Lr57* gene in the introgression lines TA5602 and TA5601.

An F₂ mapping population was previously created by crossing the rust resistant TA5602 line with the rust susceptible disomic addition line TA7659 [DS 5M^g (5D)]. Due to the wild type *Ph1* promoting homologous pairing, the 5M^g segment of TA5602 will only pair with the corresponding segment of TA7659 but still produce fertile progeny (Figure chapter 1.3). This similarly allows 5M^g recombination along the introgression, making it possible to map *Lr57* and *Yr40*.

The F_{2:3} progeny of this cross were screened for leaf rust using the isolate PTRUS55 and stripe rust using the isolate AR90-1 while seedlings. Hypersensitive responses were occasionally seen in the parental lines, in agreement with past studies, indicating *Lr57* and *Yr40* could potentially be nucleotide binding leucine rich repeat

(NLR) type genes. The rust screenings also indicated that *Lr57* and *Yr40* are separate genes, as leaf and stripe rust resistance was not always inherited together. We noticed 17 breaks points in recombinant lines for *Lr57* and ~22 for *Yr40*.

Marker data indicated that *Lr57* is between 3.9 and 5.5 Mb on 5DS. In this region, between 4 and 4.3 Mb, there are three NLR genes annotated on the wheat 5D reference and one pseudo-NLR gene, making these genes strong candidates for *Lr57*. No recombination was indicated by our markers at 4, 4.2, or 4.3 Mb, preventing us from ruling out any of the NLR genes as candidates. However, through this study we have shown a way of feasibly mapping alien genes. This is a crucial step toward more easily incorporating the useful variation within wild relatives of wheat in breeding programs. Furthermore, a genetic linkage map of our mapping population as compared to the overall TA5602 introgression size shows the high resolution that can be achieved by this method.

Materials and Methods

Table 3.1. presents a detailed description of the plant materials used. The TA6675 line contains 20 pairs of wheat chromosomes and a chromosome pair 5M^g from *Ae. geniculata* replacing chromosome 5D of wheat.

Genetic Mapping Population

Resistant *Ae. geniculata* translocation line (containing *Lr57* and *Yr40* genes) TA5602 was crossed with a susceptible wheat-*Ae. geniculata* disomic addition line for

Table 3.1: A list of wheat germplasm used in this study.

Plant ID	Plant species	Type of germplasm	Ploidy	Description
TA10437	<i>Ae. geniculata</i>	Wild wheat	(2n=4x; UUMM)	Resistant to leaf, stripe, and stem rust diseases
TA1800	<i>Ae. geniculata</i>	Wild wheat	(2n=4x; UUMM)	Susceptible to leaf, stripe, and stem rust diseases
WL711	<i>T. aestivum</i>	Hexaploid wheat	(2n=6x: AABBDD)	A susceptible spring wheat cultivar
TA6675	<i>T. aestivum</i>	Hexaploid wheat	(2n=6x: AABBDD)	The TA6675 line contains 20 pairs of wheat chromosomes and chromosome pair 5M ^g from <i>Ae. geniculata</i> , substituting for chromosome 5D of wheat
TA5601	<i>T. aestivum</i>	Hexaploid wheat	(2n=6x: AABBDD)	A Wheat- <i>Ae. geniculata</i> translocation line with ~5% of 5M ^g short arm segment translocated on 5D
TA5602	<i>T. aestivum</i>	Hexaploid wheat	(2n=6x: AABBDD)	A Wheat- <i>Ae. geniculata</i> translocation line with ~20% of 5M ^g short arm segment translocated on 5D
TA7659	<i>T. aestivum</i>	Hexaploid wheat	(2n=6x:44 AABBDD+5M)	A disomic addition line where 5M ^g of a susceptible <i>Ae. geniculata</i> is added to Chinese Spring wheat

chromosome 5M^g (TA7659). From this cross, 15 F₁ plants were grown. F₂ seeds from 6 of the F₁ plants were grown to develop an F₂ population that was screened for leaf rust response. A further F_{2:3} population included 116 families screened again for leaf rust response and 112 lines screened for stripe rust response. These populations were used in this study for genetic mapping of *Lr57*.

Previous Cytological Evaluations: Cytological Studies

For meiotic analysis, spikes of interspecific F₁ plants were fixed in Cornoy's solution (6 ethanol: 3 chloroform: 1 acetic acid) for 24 hours and transferred to 70% ethanol. Anthers at various stages of meiotic division-I were squashed in 2% acetocarmine and the pollen mother cells (PMCs) were scored for chromosomal pairing in all the crosses. Photographs were taken with a digital camera (Canon PC1049, No. 6934108049). Pollen stainability was measured by staining the pollen grains after squashing the anthers in Iodine-Potassium Iodide solution (I-KI).

Generating 5M^g specific PCR Based Markers and Genotyping of a Mapping Panel

The wheat 5D reference was used as the starting point for 5M^g-specific primer design, because *Ae. geniculata* chromosome 5M^g shares high synteny and collinearity with 5D, but 5M^g has no complete reference sequence itself (Tiwari et al., 2015). Our approach was to use 5M^g sequence with high 5DS homology as an approximation for identifying 5M^g sequences localized to our region of interest. Wheat 5DS gene sequences spanning 0 to 40 Mb were selected from the HighConfidenceGenesv1.1 IWGSC JBrowse annotation track to serve as queries for command line BLAST against

a 5M^S scaffold derived from the rust-resistant *Ae. geniculata* accession TA10437 (Camacho et al., 2009.; Consortium (IWGSC) et al., 2018; Tiwari et al., 2015). The wheat 5D and 5M^S assembly matches as well as high-scoring BLAST sequence matches from wheat 5A and 5B were aligned using GSP to design 5M-specific markers (Y. Wang et al., 2016). One marker was found at approximately 4 Mb by using a 90K SNP array and identifying one SNP that showed a polymorphic allele between *Ae. geniculata* accessions TA10437 and TA2899 (S. Wang et al., 2014). To verify 5M^S-specificity, primers were tested in duplicate on TA10437, the introgression lines TA5602 and TA5601, WL711, and TA7659. Primers that amplified WL711 or failed to amplify TA10437 were excluded from further testing. TA7659 contains a rust-susceptible 5M^S chromosome, and so both amplification and failure to produce amplicons could be expected. Moving from the telomeric region to the centromeric region, it was predicted primers would change from amplifying both introgression lines and TA10437 to only amplifying TA10437 and TA5601, indicating the end of the TA5602 5M^S translocation.

To distinguish PCR amplification failure from dominant marker polymorphism between the introgression lines and WL711, primers of interest were also multiplexed with internal control primers during PCR that could amplify uniformly across all lines and be distinguished by size on an agarose gel.

If TA7659 and WL711 were not amplified by the 5M^S primers, those primers could serve as dominant markers for mapping *Lr57* and *Yr40* in the mapping population developed from the cross between TA7659 and TA5601. If a primer pair did amplify

TA7659, it was sequenced along with TA5602, TA5601 and TA10437, and the sequences were then aligned on Clustal Omega to identify any sequence polymorphism, including SNPs, deletions, etc., which were then verified by referring back to the *ab1* sequences (Sievers & Higgins, 2014). Similarly, these sequence markers were then also used for mapping *Lr57* and *Yr40* in our mapping population.

Screening Tests for Leaf Rust Resistance

All the leaf and stripe rust phenotyping tests of the mapping population of TA5602 X TA7659 took place in Kansas, both at the F₂ stage and F_{2:3} stage for reaction against leaf and stripe rust. The 172 F₂ individuals along with parental lines TA7659, TA5601, TA5602, TA10437, and WL711 were first screened with leaf rust isolate MCDL at the two-leaf leaf seedling stage. In 2019, 116 of the F_{2:3} families, with 10 plants grown per family, were phenotyped for leaf rust reactions with the isolate PRTUS55 at seedling stage following Kallia et al., 2015 (Kalia, 2015). All plant material tested for resistance to leaf rust was grown in a 1:1 vermiculite-soil mixture in pots. Ten seeds per parental accessions and control lines were planted between two pots and grown in a greenhouse with temperature maintained at 20±3°C. WL711 and translocation line TA5602 were used as controls. Urediniospores of leaf rust cultured stored at -80°C were heat shocked at 42°C for 6 minutes before inoculation. Ten-day old seedlings of the F₂ population and controls were inoculated by spraying the seedlings with a suspension of urediniospores in Soltrol 170 isoparaffin light mineral oil (Chevron Phillips Chemical Company LLC, The Woodlands, TX). After the evaporation of the mineral oil, the inoculated seedlings were incubated in a dew chamber at 20±2°C for 24 hours. After 14 days of the inoculation the phenotypic data

on disease spread was recorded using the 0-4 Stakman scale as described in Roelfs (Table 3.2) (Roelfs, 1992)

Table 3.2. Description of susceptible and resistant leaf rust scores as per Stakman scale. The disease severity can also be modified to one end or another by adding “+” or “-”.

Stakman score	Description	Phenotype
0	No uredinia or other macroscopic sign of infection	immune response
;	Hypersensitive necrotic or chlorotic flecks without uredinia	highly resistant
1	Small uredinia surrounded by necrosis	resistant
2	Small to medium sized uredinia surrounded by necrosis	moderately resistant
3	Medium to large sized uredinia with or without chlorosis	Susceptible
4	Large sized uredinia without chlorosis	highly susceptible

Stripe Rust Phenotyping

All stripe rust disease evaluation experiments were conducted in the greenhouse and growth chamber under controlled conditions. For seedling testing, all the entries i.e. TA5601, TA5602, TA7659, mutant 125-2, mutant 125-4, mutant 666-5 and WL711 were grown in a 1:1 vermiculite: soil mixture in 4.5-cm-diameter pots. Five seeds per entry were planted in each of two pots and with the temperature maintained at $20\pm3^{\circ}\text{C}$. To evaluate for stripe rust, all the entries were inoculated with stripe rust race AR90-1 in the greenhouse. Urediniospores of stripe rust race stored at -80°C were heat shocked at 42°C for 6 minutes before inoculation. Ten-day old seedlings of all the entries were

inoculated by spraying the seedlings with a suspension of urediniospores in Soltrol 170 isoparaffin light mineral oil (Chevron Phillips Chemical Company LLC, The Woodlands, TX). The oil was allowed to evaporate, and the inoculated seedlings were incubated for 16-20 hours in a dew chamber at $12\pm 2^{\circ}\text{C}$. After this, the seedlings were transferred to a growth chamber, with temperature settings of 12°C at night and 15°C at day and a 16-hour photoperiod. Infection types (ITs) were recorded 18-21 days post inoculation, using a 0 - 9 scale (McNeal et al. 1971). Plants with infection types from 0-2 were considered resistant. Plants scoring between 3-5 were considered moderately resistant while those scoring between 6 and 9 ranged from susceptible to highly susceptible.

The same procedures were followed to phenotype the 112 $F_{2:3}$ families derived from the cross of TA7659 X TA5602 for stripe rust response. Ten seeds of each $F_{2:3}$ family as well as parental lines were grown in pots and were inoculated with stripe rust race AR90-1 at the two-leaf stage. Data recording was done according to the method described above.

Table 3.3. Description of susceptible and resistant stripe rust scores as per the McNeal scale. The disease severity can also be modified to one end or another by adding “+” or “-”. Adapted from Roelfs (Roelfs, 1992).

McNeal scale	Description	Phenotype
0	No uredinia or other macroscopic sign of infection	immune
1	No uredinia but necrotic flecks	very resistant
2	No uredinia but more necrotic flecks in stripes	resistant
3	Few uredinia with necrosis/chlorosis	moderately Resistant
4	Light uredinia stripes with necrosis/chlorosis stripes	light moderate
5	Some uredinia stripes with necrotic/chlorotic stripes	moderate susceptible
6	Moderate uredinia stripes with necrotic/chlorotic stripes	high moderate
7	Many uredinia stripes with necrotic/chlorotic stripes	moderate susceptible
8	Many uredinia stripes with chlorosis	susceptible
9	Many uredinia stripes without chlorosis	very susceptible

DNA Extractions, PCR, and Sequencing:

Parental leaf tissue was collected on ice from plants at the seedling stage and then stored at -80°C before extraction. DNA was extracted using the Kingfisher Flex DNA extraction robot with the Biosprint 96 DNA extraction kit (Qiagen, Valencia, CA, United States) and then quantified. DNA samples were diluted to 25 ng/μL. For the mapping population, a similar approach was used, except leaf tissue was collected in

bulk from 10 individual F₃ plants from the same family for each of the families in the segregating population.

PCR was carried out using Bioline MyTaq PCR kits (Bioline, Taunton, MA, United States) in 10 µL reactions. Touchdown PCR profiles were set up in BioRad 100 thermocyclers (BioRad, Hercules, CA, United States) with these conditions: 95°C- 5 min for the first denaturation step, six cycles of 95°C- 1 min, 63°C- 1 min with a decrease of 1°C per cycle, and 72°C- 2 min, followed by 25 cycles of 95°C- 1 min, 52°C-56°C for annealing depending on primer melting temperatures-1 min, and a final extension of 72°C- 7 min.

Half of the PCR product volume would be visualized on 1% agarose gels, while the second half would be reserved for sequencing, to confirm amplification specificity to the predicted 5M^g region. Sanger sequencing was done on the ABI3739xl (Applied Biosystems, Foster City, CA, United States). FASTA files were bulk extracted from the ab1 files using Biopython and then compared to the original 5M^g scaffold.

Mapping of *Lr57* and *Yr40* Genes:

From the 172 leaf rust phenotyped F₂ families, 116 leaf phenotyped F_{2:3} families, and 112 stripe rust phenotyped F_{2:3} families, each derived from the cross of TA5602 (resistant source for both *Lr57* and *Yr40*) with the disomic addition line TA7659 (susceptible for both the genes), were used to phenotype for *Lr57* and *Yr40*. The mapping population was genotyped using 5M^g specific markers using lab-based PCR assays. Genetic mapping for the *Lr57* and *Yr40* genes was performed by combining genotypic and phenotypic datasets.

Genotyping and Phenotyping of the F₂ Mapping Population and Evaluation of the F₂ Derived F_{2:3} Lines.

Estimation of the Size of the 5M^g Segment in TA5601 and TA5602:

To estimate the size of the 5M^g translocation in the translocation lines TA5601 and TA5602, PCR based markers were developed using comparative genome mapping of highly collinear 5D sequences of wheat with chromosome 5M^g of *Ae. geniculata*. PCR based 5M^g gene specific markers were assayed at every Mega base up to 60 Mb to identify breakpoints on 5M^g.5D translocations

Estimation of Mapping Resolution of the F_{2:3} Population with Respect to the 5M^g Region:

Markers that were shown to be polymorphic between TA7659 and TA5602 were then used to genotype the TA7659xTA5602 mapping population. Because DNA of the mapping population was composed of bulked tissue from individual plants from the same family, both the families showing resistance and those with heterozygous phenotypes would appear the same on a gel given a dominant 5M^g resistant marker. Heterozygous and homozygous resistant alleles could be differentiated accurately with sequence-based markers. Genotyping data was then used to create a genetic map using R/qtl/, which utilizes hidden Markov models to better account for the genotyping information of dominant markers, and centimorgans were calculated with the Kosambi function (Broman et al., 2003; Kosambi, 1943). Creation of a genetic linkage map

helped to verify the estimated 5M^g amplicon distributions across 5M^g and collinearity with 5D.

RESULTS

Evaluation of 5M^g Translocation Size in the Resistant Introgression Lines TA5601 and TA5602

We developed 24 5M^g specific markers using the high-quality resources available to us. Previous results published from our group indicated that 5M^g and 5D have high synteny and collinearity regarding gene order. This finding means we can use the order of wheat 5D genes to calculate the size of 5M^g translocations in TA5601 and TA5602. Our 24 5M^g specific markers only amplify in 5M^g but not in CS (reference wheat) or WL711 (parental line for the two introgression lines). These markers spanned 1.0 Mb to 100.0 Mb, as per the 5D chromosome of reference wheat.

Confirmation of Accurate Marker Design

Multiple markers were developed with 5M^g -specificity to characterize the TA5602 5M^g introgression size by looking for dominant marker polymorphism between TA5602, TA5601, TA10437 and WL711. To confirm that any lack of amplified product from TA5602 was not due to PCR failure, other primers were used as internal controls through multiplexing. These internal control primers amplified products of different sizes, chosen in order to be distinguishable from the PCR

product of interest (Figure 3.1). This confirmed that there were no issues with PCR failure amplifying TA5602 and true dominant markers could be found.



Figure 3.1: 5M^g dominant markers specific to resistant *Ae. geniculata*. A marker amplifying a segment of leaf rust resistant *Ae. geniculata* at 4 Mb (upper band) multiplex with an internal control (lower band) to screen for PCR failure. Only the introgression lines TA5602, TA5601, and the parental line TA10437 show amplification due to the 4-Mb 5M^g marker. The rust-susceptible *Ae. geniculata* line TA1800, TA7659, and Chinese Spring (CS) were also included to further show resistant 5M^g specificity. 5M^g-specific marker 4.3 Mb (lower band) multiplex with internal control (upper band). The rust susceptible wheat lines WL711, disomic addition line TA7659, CS, and TA1800 all fail to amplify the 5M^g resistant *Ae. geniculata* segment (upper band).

The Introgression Size of TA5602

Markers were developed spanning the length of 5M^g, first going up to approximately 17 Mb (Table 3.4). From our data, markers at a location greater than or equal to 9.548 Mb consistently amplified TA5601 5M^g, but not WL711 or TA5602. These markers also consistently amplify TA10437 and generally TA7659, which both have full length 5M^g chromosomes (Figure 3.2). Failure to amplify TA7659 is an indication that a marker is specific to rust resistant 5M^g. It was also consistently shown that any marker amplifying segments of 5M^g at locations less than or equal to

8.9 Mb could amplify both TA5602 and TA5601. In summary, TA5602 5M^g

introgression constitutes somewhere between 8.9 and 9.548 Mb of 5DS.

Table 3.4: List of primers designed to characterize the 5M^g introgression size of TA5602. The approximate amplification location is based on the 5M^g introgression target location in relation to wheat chromosome 5D. All primers that amplify TA5602 and TA5601 also amplify TA10437. None of the primers amplify WL711, as confirmed by internal controls, suggesting 5M^g-specificity. '+' and '-' indicate presence or absence of amplification.

Position on 5D	Amplifies WL711	Amplifies TA5601	Amplifies TA5602
5Mg_17Mb	-	+	-
5Mg_9.9Mb	-	+	-
5Mg_9.7Mb	-	+	-
5Mg_9.548Mb	-	+	-
5Mg_8.9Mb	-	+	+
5Mg_8Mb	-	+	+
5Mg_7.8Mb	-	+	+
5Mg_7.4Mb	-	+	+
5Mg_6Mb	-	+	+
5Mg_5.6Mb	-	+	+
5Mg_5.5Mb	-	+	+
5Mg_4.3Mb	-	+	+
5Mg_4.2Mb	-	+	+
5Mg_4Mb	-	+	+
5Mg_3.9Mb	-	+	+
5Mg_3.6Mb	-	+	+
5Mg_1.79Mb	-	+	+
5Mg_1.5Mb	-	+	+

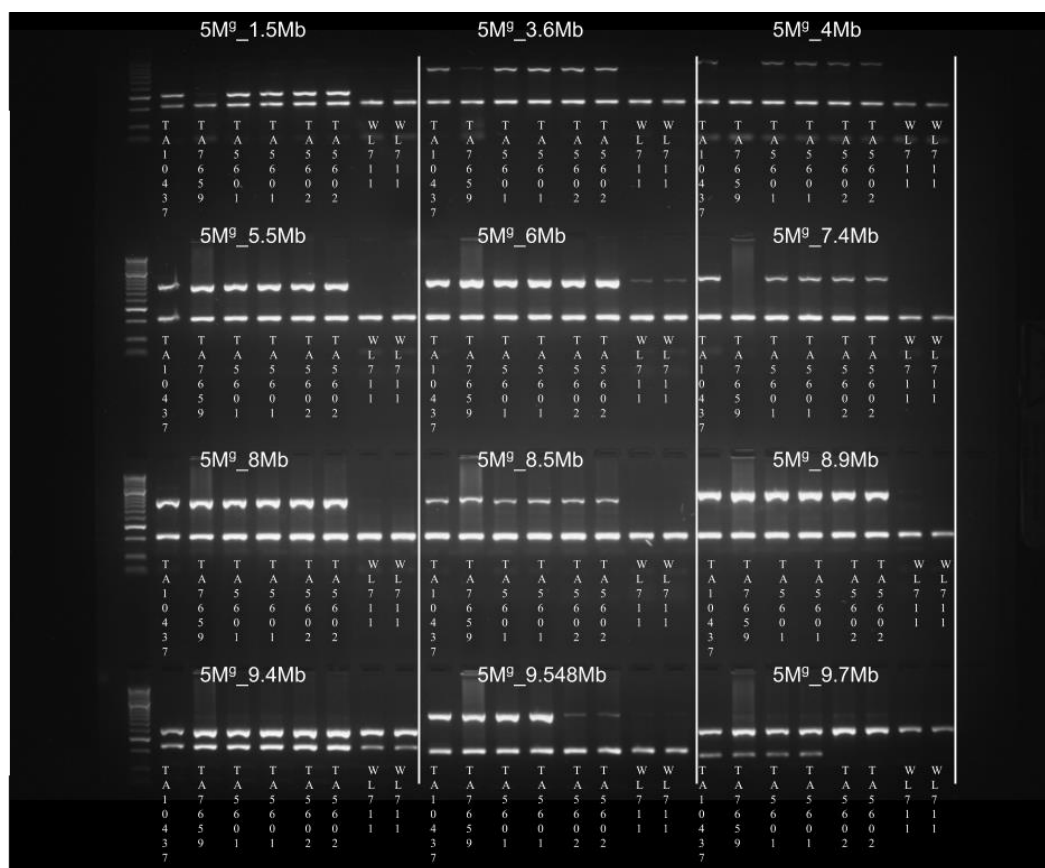


Figure 3.2: Characterizing the 5M^g introgression size of TA5602 with 5M^g-specific primers. Markers before 8.9 Mb consistently produce amplicons in both TA5602, TA5601 and TA10437. Amplicons after 9.548 Mb consistently fail to produce amplicons in TA5602 but do produce amplicons in TA5601 and TA10437. This indicates the introgression size of TA5602 is between 8.9 and 9.548 Mb on 5DS. The 9.4 Mb showed potentially non-specific binding to wheat 5D.

The Introgression size of TA5601

Additional primers were also designed to characterize the introgression size of TA5601 using the same method of designing 5M^g -specific markers for TA5602. Primers were designed continuing from 17 Mb up to 100Mb and again tested on TA5601, TA10437, TA7659, and WL711 (Table 3.5). TA5602 was also included in these primer tests and failed to be amplified as expected given its smaller introgression size. From these results, we were able to show that TA5601 and TA10437 continuously both amplified up until 50 Mb. After Mb, only TA10437 and TA7659 can be amplified. This confirms the introgression size of TA5601 is between 50 and 60 Mb (Figure 3.3).

Table 3.5: List of primers designed to characterize the 5M^g introgression size of TA5601. The approximate amplification location is based on the 5M^g introgression target location in relation to wheat chromosome 5D. All primers that amplify TA5601 also amplify TA10437. None of the primers amplify WL711, as confirmed by internal controls, suggesting 5M^g-specificity. ‘+’ and ‘-’ indicate presence or absence of amplification.

Position on 5D	Amplifies WL711	Amplifies TA10437	Amplifies TA5601
5Mg_60Mb	-	+	-
5Mg_50Mb	-	+	+
5Mg_45Mb	-	+	+
5Mg_35Mb	-	+	+

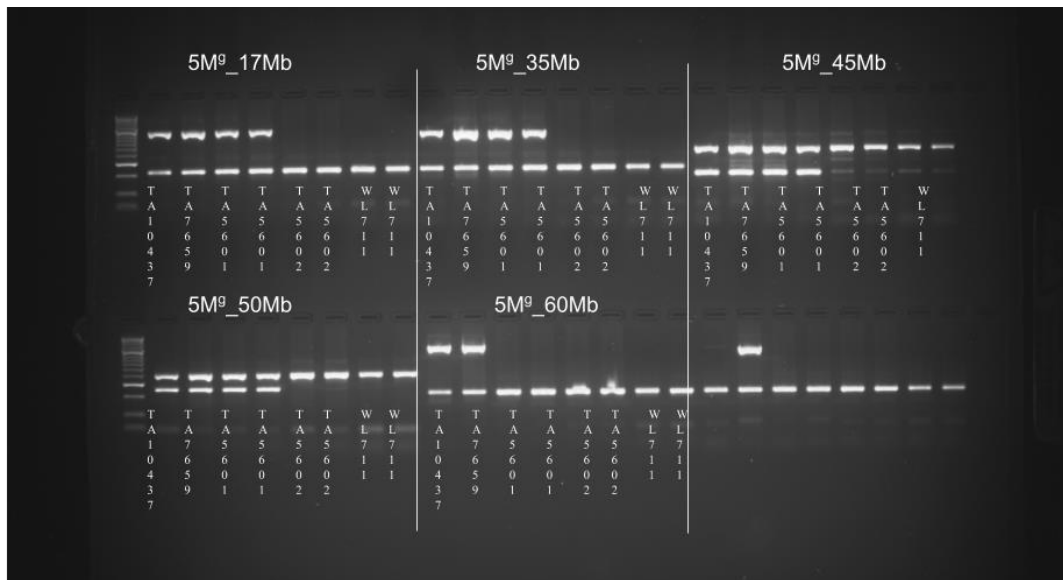


Figure 3.3 Characterizing the 5M^g introgression size of TA5601 with 5M^g-specific primers. Markers before 50 Mb consistently produce amplicons in both TA5601 and TA10437. Amplicons at 60 Mb can be produced by TA7659 and TA10437, but not TA5601, suggesting this is past the introgression site. This means the TA56015M^g introgression size is between 50 and 60 Mb.

Unique Mapping Scheme for Alien Germplasm

Estimation of the actual fragment lengths of the translocation lines provided very useful information to perform experiments precisely mapping the *Lr57* and *Yr40* genes, as they are now confirmed to be in a very small physical interval of 9.548 Mb. In similar work, Bansal et al. reported localization of two important genes *Lr76* and *Yr70* from the syntenic region of *Ae. umbellulata*. Interestingly they also reported mapping of these two genes in the terminal region of 5D.5U translocation lines. Bansal et al tried to reduce this interval by developing a large F₅ population of 1404 recombinant inbred lines, but the 9.47 interval was not reduced any further.

To establish an efficient approach to reduce this interval, we developed a high-resolution mapping population by crossing TA5602 (3.5% of 5M^g) with TA7659 (full chromosome 5M^g), so that pairing can be induced between 5M^g chromosomes from resistant and susceptible introgressions lines differing for the presence or absence of the *Lr57* and *Yr40* genes. To precisely map 5M^g specific genes, we crossed the Wheat-*Aegilops geniculata* translocation line TA5602 (resistant for leaf and stripe rust diseases) with the leaf and stripe rust susceptible wheat-*Aegilops geniculata* disomic addition line for chromosome 5M^g TA7659. A description of the germplasm is provided in Table 3.1. The rationale behind these crosses were to allow 5M^g resistant and susceptible chromosomes to pair and for the subsequent recombination events between these chromosomal regions to enable us to map resistant genes by assaying them with polymorphic markers. Figure 3.4 explains the crossing scheme for this work.

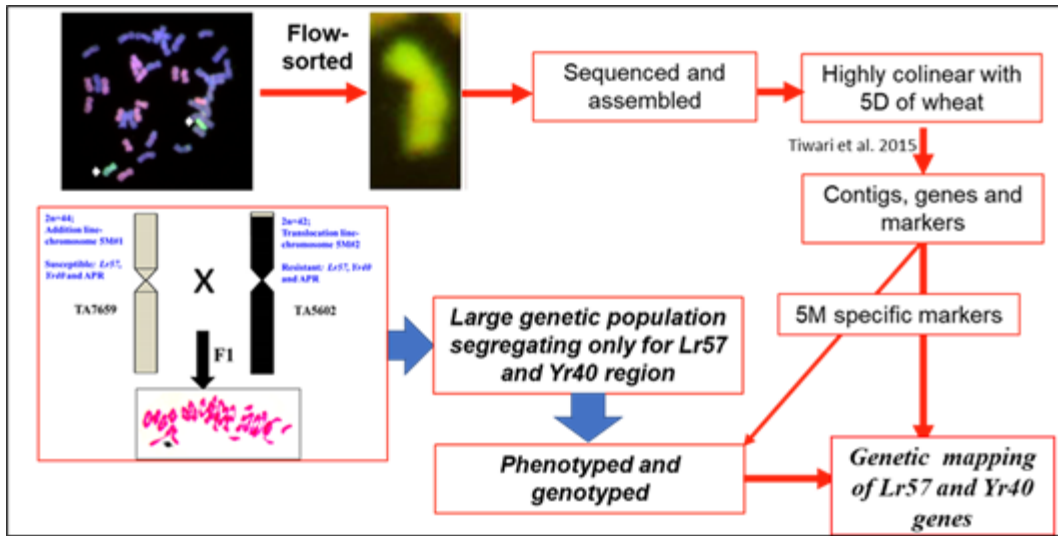


Figure 3.4. Schematic presentation of the 5M^S genomic resources and approach used in this study to perform high-resolution genetic mapping of 5M^S fragments coming from two different accessions showing differential responses against leaf rust and stripe rust pathotypes.

Cytological Analysis

Since in this study a cross was made between two alien introgression lines, to confirm the pairing behavior, we performed cytological analysis of the F₁ plants. Developing spikes were fixed at the meiosis and chromosome pairing data was recorded (Figure 3.5).

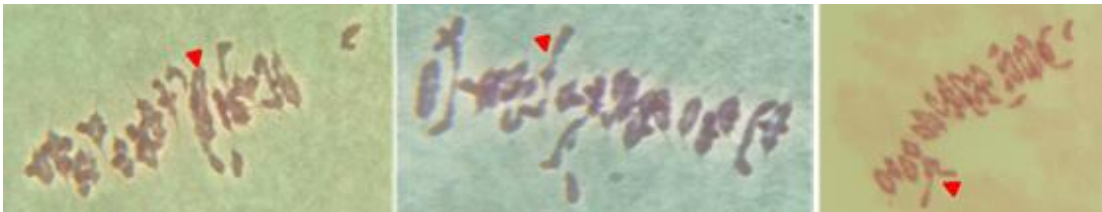


Figure 3.5. Chromosome pairing data suggesting development of a trivalent between the tip of the 5M^S resistant chromosome and full 5M^S susceptible chromosome.

Normally, in bread wheat, cytological analysis at meiosis shows 21 bivalents (in a ring-like structure showing pairing of the chromosomes). In this cross, we were expecting chromosome pairing between the tips of the 5M^g fragment of TA5602 and the full chromosome of 5M^g from TA7659. Since the long arm of the 5D (TA5602) and 5M^g (TA7659) will not pair, we expected to see a trivalent in the F₁ plants. Analysis of about 500 PMC clearly indicated the presence of a trivalent in all the PMCs scored with 1-2 expected univalent and 19-20 bivalents. Trivalent presence clearly indicated 5M^g pairing between TA7659 and TA5602. In the F₂ population derived from the plants showing a trivalent, recombination events are expected between 5M^g resistant and susceptible chromosomes. We used the F₂ seeds from F₁ plants (2013-91-11; trivalent were confirmed in cytological analysis). A population of 148 F₂ individuals were grown and evaluated to test for applicability for mapping purposes.

Phenotyping leaf and Stripe rust

Figure 3.6 shows phenotyping results of WL711 TA5601, TA5602, TA7659 for resistance against the leaf rust isolate PRTUS55. Interestingly, TA5601 showed consistent scores of “;”, indicating hypersensitive flecks but no uredinia spores, as per the Stackman scale (Table 3.2) (Roelfs, 1992) while TA5602 received scores between 1 and 1+. This suggested that TA5601 contained a second gene for resistance, which TA5602 lacks. TA5602 EMS mutants consistently showed scores of between 7-8. Susceptible *Ae. geniculata* 5M^g addition line TA7659 received consistent scores of “3” and WL711 received scores between 3+ and 4.

One hundred and sixty-two plants of the F₂ population were phenotyped, out of which 122 were found to be resistant and 40 were found to be susceptible, fitting the genetic ratio of 3:1. This indicates *Lr57* is a single, dominant gene in TA5602.

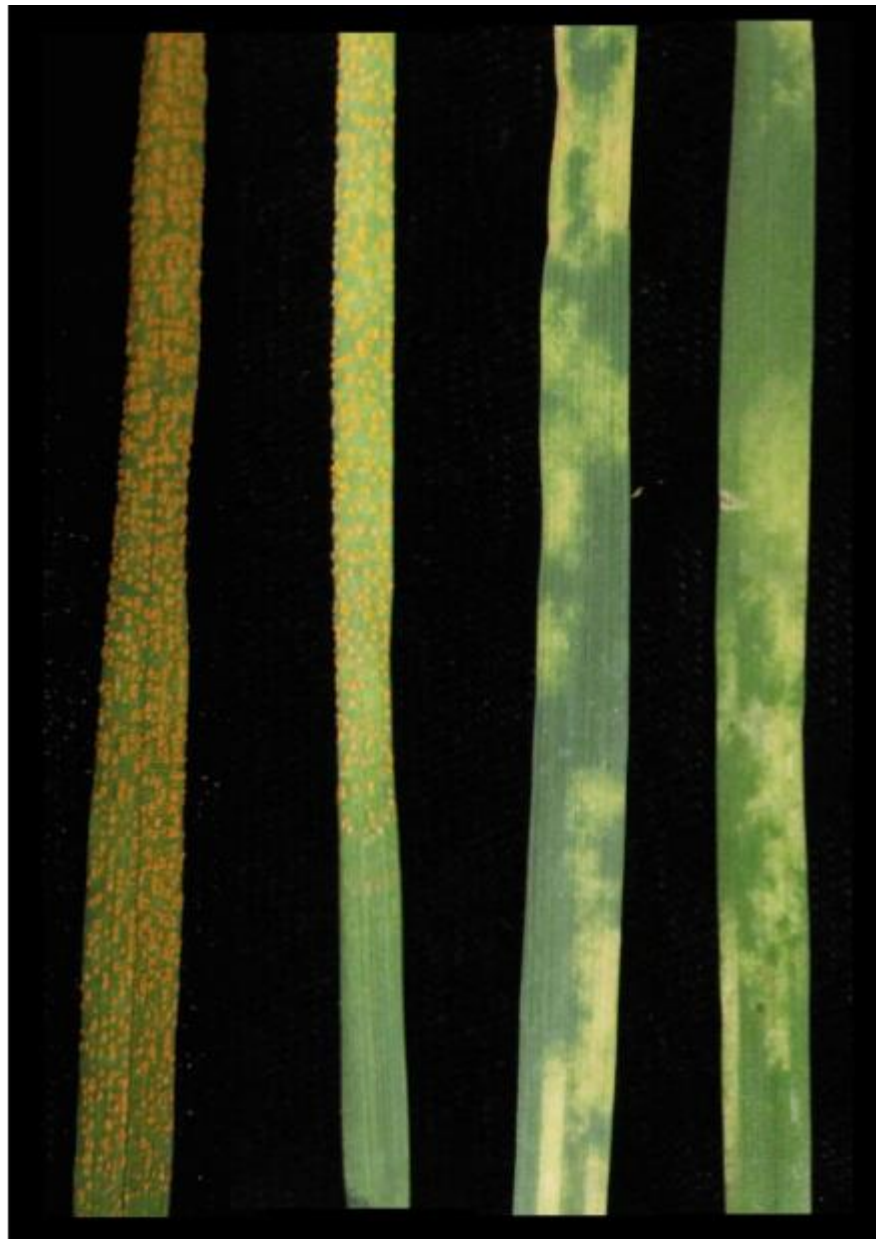
One hundred and sixteen F_{2:3} families were tested further for leaf rust response and 32 were found to be homozygous resistant, 55 were found to be heterozygous, and 29 were found to be homozygous susceptible, fitting an expected 1:2:1 segregating ratio.

Stripe rust inoculation at the seedling stage was done for the parental set and one hundred and twelve of the F_{2:3} families. The susceptible parental wheat line WL711 showed consistent high susceptibility scores of between 7-8 on the stripe rust McNeal scale after inoculation with stripe rust isolate AR90-1 (McNeal et al., 1971). TA7659 similarly showed consistent scores between 7 and 8. Interestingly, introgression line TA5602 consistently showed resistant scores between 1-2, while TA5601, which contains a larger amount of *Ae. geniculata* chromatin, showed a consistent score of 0 (Figure 3.7), indicating the possibility of additional resistance gene in TA5601.

In the 112 F_{2:3} families, 22 families were found to be homozygous resistant, 58 were found to be heterozygous, and 32 were found to be homozygous susceptible.



Figure 3.6: Leaf rust phenotyping. Leaf rust inoculations during the seedling stage with the isolate PRTUS55. Inoculations pictured include the disomic addition line TA7659 [DS5M^g (5D)]. The introgression wheat-*Ae. geniculata* introgression lines TA5602, and TA5601. TA5601 shows stronger resistance than TA5602. Also shown are examples of rust susceptible and rust resistant F_{2:3} individuals.



WL711 TA7659 TA5602 TA5601

Figure 3.7: Stripe rust results. Stripe rust isolate AR90-1 inoculations during the seedling stage. Inoculations on the wheat line WL711 and the disomic addition line TA7659 [DS5M^g (5D)]. The introgression wheat-*Ae. geniculata* introgression lines TA5602, and TA5601. TA5601 and TA5602 both show high stripe rust resistance.

Initial *Lr57* and *Yr40* Analysis:

From the phenotyping data, multiple F_{2:3} families were found that showed discrepancies between leaf and stripe rust responses (Table 3.6). If *Lr57* and *Yr40* were

one gene showing pleiotropic effects, we would expect both leaf and stripe rust resistance and susceptibility to be inherited together. From our data, we expect *Lr57* and *Yr40* to be separate genes. Given we had access to an RNAseq data set comparing WL711 and TA5602 response to leaf rust, we decided to focus primarily on mapping *Lr57*.

Table 3.6: A subsection of lines derived from a cross between TA5602 and TA7659 that were phenotyped both with the leaf rust isolate PRTUS55 and the stripe rust isolate AR90-1. Given the differences in phenotype scores for leaf and stripe rust, there is evidence that *Lr57* and *Yr40* are separate genes. Leaf rust scores are between 1 and 4, with “;” indicating hypersensitive fleck formation. The individual leaf and stripe rust scores are based on the Stackman scale and McNeal scale respectively. “A” represents resistant lines, “B” susceptible lines, and “H” are heterozygous lines.

F2:3 Family	Leaf Rust	Stripe Rust
25	H	S
30	R	S
31	H	S
35	H	S
36	H	S
154	H	S
57	H	S
69	H	R
84	R	H
128	R	H
129	S	H
139	R	H
148	S	H
153	H	R

The *Lr57* Candidate Region

Ten of the markers used to characterize the TA5602 introgression size were used as part of mapping *Lr57* (Table 3.7 and Table 3.8).

Table 3.7: The primers used to map Lr57. +/- refers to dominant, gel-based markers (presence/absence on resistant/susceptible lines).	
Marker Name	Polymorphism
1.5Mb	+/-
1.79Mb	+/-
3.6Mb	+/-
3.9Mb	+/-
4Mb	+/-
4.2Mb	+/-
4.3Mb	+/-
5.5Mb	C/T
5.6Mb	C/T
6Mb	G/T

From our genotyping and phenotyping data, *Lr57* was determined to be between approximately 3.6 and 5.5 Mb on 5M^g and closely linked to two markers in particular at 4.2 and 4.3 Mb. From the wheat 5D reference annotation, this region contains three NLR genes (TraesCS5D02G005300, TraesCS5D02G005400, and TraesCS5D02G005600) and a pseudo-NLR (TraesCS5D02G005416) gene between 4.0 and 4.3 Mb, with the next closest NLR genes being located at roughly 10 and 1.8 Mb. Because of this region of interest, observations of seedling-stage resistance, and *Lr57* showing hypersensitive responses against certain rust isolates like PNMQ (Kuraparthi et al., 2007), these NLR genes became the main candidates of interest.

Two more markers were developed at 3.9 and 4 Mb, with the 4Mb marker being found by using BLAST of

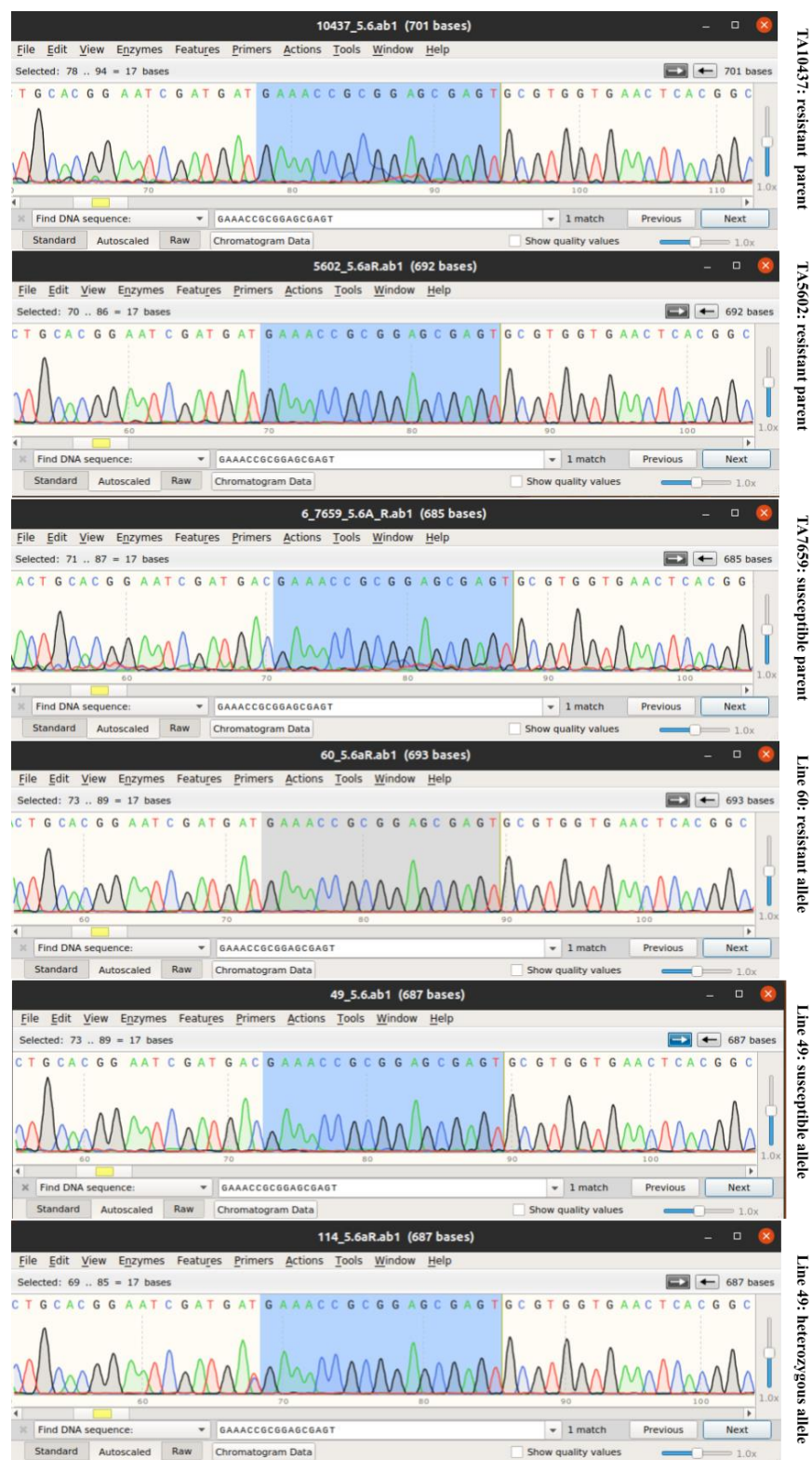


Figure 3.8: An example of a SNP based marker (targeting 5M^s that has homology with the wheat gene TraesCS5D02G006800). T/C SNP.

the 5D TraesCS5D02G005300 gene sequence against a 90K SNP array and accessing the resulting alleles for differences between TA10437 and TA2899 (Shichen Wang et al., 2014). From these markers, we could see that any *Lr57* candidate gene clearly would be between at least 3.9 and 5.5 Mb (Table 3.9). No differences were seen between the markers genotype data using the 4, 4.2 and 4.3 Mb markers, preventing us from ruling out any of the NLR candidate genes in this region.

Table 3.8: A subsection of marker scores as tested on the mapping population derived from a cross between leaf rust susceptible disomic addition line TA7659 and leaf rust resistant introgression line TA5602, as phenotyped at the F_{2:3} stage by leaf rust isolate PRTUS55. Each marker before 5.5 Mb are based on gel-polymorphism. Scores of ‘+’ refer to the TA5602 resistant genotype and a score of ‘-’ refers to the TA7659 susceptible genotype. Line 114 shows that the candidate region will be after 3.9 Mb while lines 49 and 156 are both examples of the candidate region being before 5.5 Mb.

5D Name	TraesCS5D02G001000	TraesCS5D02G001400	TraesCS5D02G004600	TraesCS5D02G005200	TraesCS5D02G005300	TraesCS5D02G005400	TraesCS5D02G005500	TraesCS5D02G006600	TraesCS5D02G006800	TraesCS5D02G007200	F2:3
5D Location (mB)	1.5	1.79	3.6	3.9	4	4.26	4.309	5.5 mB	5.6 mB	6 mB	Leaf Rust
49	+	+	+	n/a	+	+	+	-	-	-	+
17	-	-	-	+	+	+	+	+	-	-	+
44	-	-	-	+	+	+	+	+	+	+	+
142	-	-	-	+	+	+	+	+	+	+	+
114	+	+	+	+	-	-	-	+	+	+	-
21	+	+	+	-	-	-	-	+	n/a	+	-
156	+	+	+	-	-	-	-	-	-	+	-

These 10 polymorphic markers were used to genotype our mapping population with 172 F₂ plants, 116 F_{2:3} plants phenotyped for leaf rust response, and 112 F_{2:3} plants phenotyped for stripe rust response. We were able to develop a tentative map of the 9.548 Mb TA5602 introgression region. Genetic mapping of this region resulted in a

genetic map of the first 6 Mb of this region and yielded a total map length of 11.4 centimorgans (Figure 3.10). The map order of our markers was in perfect collinearity with a reference chromosome of 5D of bread wheat.

Interestingly, Bansal et al., (2020) recently published work on a similar project, attempting to map the *Lr76* and *Yr70* genes using 1404 F₅ recombinant inbred lines derived from a wheat-*Ae. umbellulata* translocation line pau16057, where 5DS was replaced with 5U, was crossed with WL711. In their project, 27 5U-KASP markers identified a single 9.47 Mb non-recombinant 5U block on 5D, which is not surprising given the *Ph1* loci will make homoeologous recombination between 5D and 5U unlikely. Due to the lack of recombination, further narrowing of their candidate region was not possible. This shows the power of our mapping approach, where fewer markers (n=10) on a smaller population (n=116) allowed us to narrow down the *Lr57* candidate region to a small interval between 3.9 and 5.5 Mb.

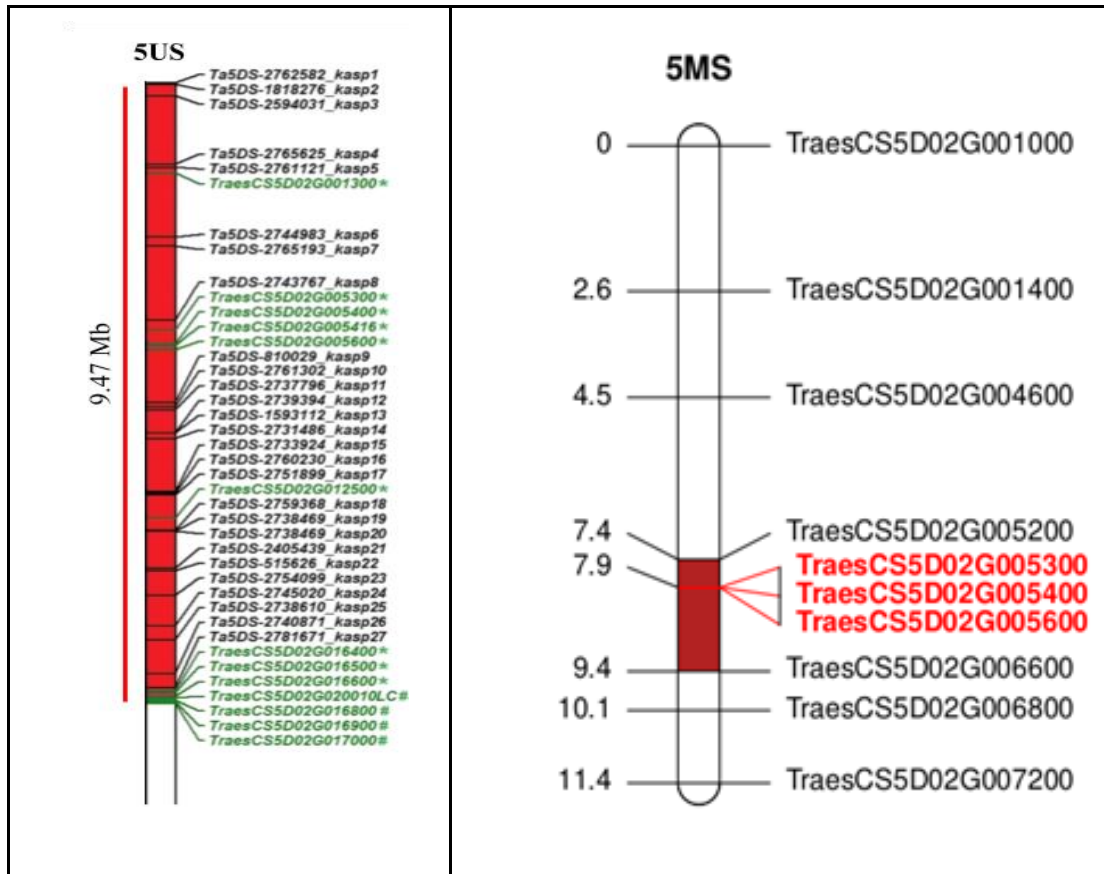


Figure 3.9: **Left:** the map of the 5U-5D short arm of the introgression line pau16057 as developed by Bansal et al. 2020. In red is the polymorphic region between pau16057 and 5D as identified with 27 KASP markers on a 1404 F₅ population. However, no recombination was identified in this region in the F₅ population. **Right:** Genetic map of the tip of chromosome 5M^S based on the mapping population derived from TA5602 x TA7659. A leaf rust mapping population of 172 F₂ plants and 116 F_{2:3} plants were developed for genotyping and phenotyping. In red is the candidate *Lr57* region with NLR genes highlighted in red. This map was developed from screening with 10 polymorphic markers. This shows the relatively high recombination mapping power of our mapping technique.

Mapping Resolution

Availability of the 5M^S genetic map spanning a 6.0 Mb region provided us an opportunity to calculate mapping resolution of the novel mapping scheme we proposed.

Total physical length assayed in this study was 6 Mb and using 10 markers a genetic map length of 11.4 centimorgans was generated. So, map resolution of the small mapping population was estimated as 526 Kb per centimorgan (6.0 Mb/11.4 cM). This is a significant achievement considering limitations in recombination-based mapping of alien germplasm.

Discussion

It was previously hypothesized that TA5602 constituted less than 3.5% of 5DS and could be characterized as a cryptic introgression due to being “cytologically invisible” with GISH probes (Kuraparthi, 2007). To test this, we were able to first design markers confirmed to be *Ae. geniculata* 5M^g-specific and, in doing so, showed the value of comparative genomics in facilitating alien chromosome primer design. Well-characterized wheat reference sequence can be aligned against less-well-characterized wild wheat relative sequences to facilitate easy primer design.

From our marker data, we can clearly see that the TA5602 introgression size is somewhere between 8.9 and 9.548 Mb on 5DS. At 9.548 Mb or greater, markers consistently fail to amplify TA5602 but do amplify the larger 5M^g chromosome/segment of TA10437 and TA5601. We now know that wheat chromosome 5D is 566 Mb, making the TA5602 introgression constitute approximately 1.7% of 5D (Consortium (IWGSC) et al., 2018).

Similarly, we used the same methodology to characterize the introgression size of TA5601, comparing TA5601 versus TA10437 amplification. This showed the introgression size of TA5601 is between 50 and 60 Mb. During the process of

characterizing the introgression lines, multiple polymorphic markers were found between the leaf and stripe rust resistant introgression lines and TA10437 versus the leaf and stripe rust susceptible line TA7659. These markers were then used to facilitate mapping of *Lr57* and *Yr40*.

In this work, we were able to develop an alternative approach to mapping alien introgressions that involves crossing a translocation line of interest with a disomic addition line, allowing only small regions of interest to recombine for a mapping population. Using this approach, a wild type *Ph1* locus will help ensure both the fertility of progeny and that pairing and recombination only occurs between segments of homologous alien chromatin and segments of homologous wheat chromatin, allowing for the generation of an otherwise seemingly typical F₂ mapping population.

Through use of polymorphic dominant, gel-based markers as well as sequencing-based markers, we were able to determine the *Lr57* candidate region would be between 3.9 and 5.5 Mb. In this region, there are four annotated genes, three of which are NLR type genes, as based on the Chinese Spring wheat reference. This is not uncommon, as resistance genes are known to appear in clusters (Michelmore & Meyers, 1998). An NLR gene as a candidate for *Lr57* is what would be expected from the phenotyping data gathered from the parental lines, where seedlings showed resistance. We also saw TA5601 had stronger stripe rust resistance than TA5602, suggesting TA5601 had a second stripe rust resistance gene that TA5602 lacked. This

was later confirmed by a recently published paper that designated *YrAg* as a newly found *Yr* gene unique to TA5601 and absent in TA5602 (Kumar, 2020).

Interestingly, our phenotyping data showed leaf rust and stripe rust resistance were not always inherited together, ruling out the possibility that *Lr57* and *Yr40* are one gene showing pleiotropic effects.

There are two notable bottlenecks when it comes to fine mapping in tertiary gene pool members: the *Ph1* locus causing unfavorable conditions for homoeologous recombination and a lack of genomic resources. This shows that it is possible to use introgression lines crossed with disomic addition lines to overcome the challenges of the *Ph1* locus and map alien genes like *Lr57*. We have also shown we can use the better-characterized wheat reference sequence as a proxy for 5M^g, allowing us to create 5M^g-specific markers to fine map alien genes.

Chapter 4: Physical Mapping and Identification of a *Lr57*

Candidate Gene

Abstract

Initial genetic mapping yielded a 2.0 cM flanking interval for the *Lr57* locus. Since the genetic population still had mapping resolution to saturate this interval, we first tried to add more polymorphic markers to further narrow the *Lr57* candidate region. The addition of one more marker at 5.1 Mb was not able to reduce the *Lr57* candidate region-other markers likely still could. Our group's work on chromosome 5M^g clearly established very high collinearity between 5M^g and 5D chromosomes, meaning the 5M^g contigs can be ordered by alignment to the reference 5D of wheat. Additionally, it suggested that by using flanking markers of the *Lr57* region, we can perform chromosomal landing with the reference 5D and make comparison to the 5M^g assembly. Our eventual physical map of the *Lr57* region spanned a 1.1 Mb interval (3.97 to 5.00 Mb). Annotation of the 5M^g region by comparison to the 5D reference map identified 7 candidate genes in the *Lr57* physical interval. To narrow down the number of candidate genes for a final validation study, we used gene ontology studies as well as transcriptome data to rule out unrelated candidate genes. Our collaborators at PAU Ludhiana have generated RNAseq data of the 5M^g translocation line TA5602 and control wheat line WL711 after leaf rust inoculation with the isolate 77-5 (Yadav et al., 2016). This study found a higher amount of nucleotide binding and leucine rich repeat (NLR) genes expressed in TA5602, with differential expression of genes being highest 12 hours after inoculation. Since we have already delineated the *Lr57* region to

a very small physical interval, this expression data set provides a great resource to identify any *Lr57* candidate gene. Our goal was to reassemble the transcriptomes of TA5602 and WL711 while consulting the mapped *Lr57* candidate region as a reference. Contigs of 5M^g resistant scaffold developed by Tiwari et al., (2015) were used to create a physical map of the majority of the *Lr57* candidate region, spanning between 3.9 and 5 Mb and served as the reference for aligning *de novo* WL711 and TA5602 transcripts. Any transcript that was differentially expressed but was outside of our *Lr57* region was not included in the list of candidate genes. Out of 7 candidate genes from the mapped *Lr57* physical region, only AE5M2G005600 (TraesCS5D02G005600 on 5D) showed differential expression at 12-hours post-inoculation. Our results indicated that the candidate gene AE5M02G005300 (TraesCS5D02G005300 on 5D) at 4Mb was highly differentially expressed in our candidate region at 12 hours according to DESeq2 and EdgeR. However, this gene had higher expression in WL711 (susceptible wheat), not TA5602. Only one candidate gene AE5M2G005600, an NLR gene, showed higher expression in the resistant translocation line and was within our *Lr57* physical interval. A *de novo* transcriptome assembly of mock treated TA5602 did not appear to generate reads that could align to AE5M2G005600 at a significant level, suggesting this gene is induced in TA5602 after rust inoculation.

Given *Lr57* shows an NLR-type seedling resistance against leaf rust, this is a strong candidate gene and validation of this gene is underway.

Introduction

Availability of unique germplasm, a genetic mapping population, and genomic resources put us in a unique position to perform positional cloning of the *Lr57* gene. Tiwari et al, 2014, and Tiwari et al, 2015, developed pipelines and genomic resources to map and characterize 5M^g specific genes. Combining advances in NGS chemistries, availability of reference genomes of wheat, wheat's diploid and tetraploid progenitor species, and pasta wheat provides us tools and resources to make swift advances in isolating agronomically important genes from distant gene pools of wheat.

Physical Mapping and Candidate Gene Identification

Creating accurate physical maps of genetic sequences is usually a challenging task, particularly in highly repetitive and large polyploid genomes like those of the Triticeae family. The recently published Chinese Spring wheat reference, the assemblies for *Ae. tauschii* and emmer wheat, and the maize pangenome, all utilized DenovoMAGIC2 with high-coverage short Illumina reads to create their *de novo* genome assemblies (Consortium (IWGSC) et al., 2018; Lu et al., 2015; G. Zhao et al., 2017). This data was generally integrated along with numerous other techniques, such as POPSEQ and Hi-C data, to generate final reference sequences. However, proprietary software such as DenovoMAGIC2 can be something of a black box in terms of function and can be both computationally and monetarily prohibitive to use. In published, open source *de novo* genome assemblers, variations of splitting DNA reads into K-mers (smaller substrings of the DNA reads of length “K”) for de Bruijn graph-based contig assembly and then scaffold assembly is common. Overlapping raw DNA reads

themselves should not typically be used for genome assembly, particularly of genomes like that of wheat. There are several reasons this is the case. First, this can quickly turn into an NP-complete problem, meaning the assembly cannot be solved efficiently- for example, trying to assemble a million reads would require computationally evaluating a trillion pairwise alignments (Compeau et al., 2011). Furthermore, simplified approximations of this approach, such as via the “greedy” algorithm, can fail to assemble repetitive sequences (Schatz et al., 2010). Paired-read sequencing, counting K-mer frequency, and de Bruijn graphs can help resolve such issues when assembling Triticeae from short reads, as can long read sequencing like that offered by PacBio (Compeau et al., 2011).

Tiwari et al. (2015) previously assembled the NGS assembly of resistant chromosome 5M^g using SOAPdenovo, a de Bruijn graph-based assembler (Li et al., 2010; Luo et al., 2012; Tiwari et al., 2015). The 5M^g assembly resulted in ~246,000 contigs with a length ≥ 500 bp and a mean coverage depth of 13 \times . Assembled repeat-free sequences were then used for chromosome structure analysis, gene detection and construction of a virtual gene order map. About 1000 5M^g contigs were larger than 10 Kb. Overall this assembly provided a valuable resource for functional isolation of agronomically important genes from chromosome 5M^g. Due to the now available 5D reference sequence and its high level of synteny and collinearity with 5M^g, we could use comparative genomics to facilitate the construction of a 5M^g physical map in our narrow *Lr57* candidate region from the 5M^g scaffold. We were able to order the 5M^g pre-assembled contigs and orient them within the candidate region based on 5D gene orientation supplemented with our own sequencing to generate a physical map.

Transcriptome assemblies combined with reference genomes provide a powerful resource to identify candidate genes underlying a trait of interest. Recently the transcriptome landscape of wheat has been developed using a vast set of RNAseq datasets, allowing us to quantify the expression of candidate genes during developmental processes and biotic and abiotic stresses. Normally genes related to NLR gene families show differential gene expression if challenged by pathogens.

Yadav et al., 2016, conducted an RNAseq experiment on WL711 and the near isogenic line TA5602 (IL T756) (Yadav et al., 2016). In that study, the rust race *P. triticina* 77-5 was used to inoculate plants and RNA samples were collected at 0, 12, 24, 48, and 72 hours post-inoculation. TruSeq RNA library Prep Kit was used to construct libraries that were then sequenced with an Illumina HiSeq2000 to generate paired end reads. After removing low quality sequences and adapters with FastQC and cutadapter, a *de novo* assembly from the data was created using Trinity (Anders & Huber, 2010; Haas et al., 2013; Martin, 2011). At the time of the Yadav study, a complete wheat reference had not been created, necessitating the reliance on *de novo* assemblies. This study also showed 2,692 transcripts differentially expressed from TA5602, particularly nucleotide-binding and leucine-rich repeat genes (NLR genes), as classified by Blast2GO.

Previously we had broadly mapped the TA5602 5M^g introgression to a region spanning ~9.5 Mb on 5D. By crossing TA5602 with the rust susceptible disomic addition line TA7659, a leaf rust mapping population of 172 F₂ plants and 116 F_{2:3} plants were developed and phenotyped. Combining this phenotyping data with genotyping data generated from 10 polymorphic markers, we could narrow *Lr57* to a

region between 3.9 and 5.5 Mb. We also generated a genetic map spanning approximately 6 Mb, as based on the 5D reference. Our genetic map covered an area of 11.4 cM, showing the high mapping resolution of our mapping population and utility for mapping alien introgressions. Markers created based on 5D NLR genes between ~4 and ~4.3 Mb appeared to predict the F_{2:3} leaf rust response particularly well.

Given the release of the high-quality annotated wheat reference (Consortium (IWGSC) et al., 2018), we sought to reexamine the differential gene analysis between WL711 and TA5602 using the same RNAseq data generated by Yadav et al., 2016, but relying on RNA reads alignment guided by 5D. This was then shown to be ineffective. Subsequently, a second round of analysis was done with RNA reads for the 12-hour transcriptome assembled again *de novo*. These were then aligned to a 5M^g physical map that was constructed from the 5M^g scaffold contigs ordered based on 5D. Although particular gene orthologs may show presence/absence variation among homoeologous chromosomes, we do broadly know 5D and 5M^g are highly collinear and our genetic map appeared to support this as well (Tiwari et al., 2015).

Methods

Chromosome 5M^g Assembly and Physical Mapping of the *Lr57* Region:

NGS contigs generated by Tiwari et al., 2015 were used as the starting point for the construction of a putative physical map *Lr57* region. Additionally, the gap region between the markers were filled by unmapped 5M^g reads and sequences of genes from *Lr57* region using PCR-based amplification and their subsequent Sanger sequencing. Initial genetic mapping yielded a 2.0 cM flanking interval for the *Lr57* locus. Since the genetic population still had mapping resolution to saturate this interval, we tested

another polymorphic marker at 5.1 Mb. With the addition of this marker, we were not able to reduce the *Lr57* candidate region from 5.5 Mb, although it is likely still possible. Our group's work on chromosome 5M^g clearly established very high collinearity between 5M^g and 5D chromosomes, which means our 5M^g contigs can be aligned to the reference 5D of wheat. Additionally, it suggested that by using flanking markers of the *Lr57* region, we can perform chromosomal landing on the reference of 5D and comparatively, the 5M^g assembly. The physical map of the *Lr57* region spanned 1.1 Mb interval (3.97 to 5.00 Mb). Annotation of the 5M^g region and comparative analysis of the 5D reference map identified 7 candidate genes in the *Lr57* physical interval. To narrow down the number of candidate genes for final validation studies, we used gene ontology studies as well as transcriptome data to rule out unrelated candidate genes.

5D Guided Alignment of RNAseq Reads

First the RNA paired end reads were uploaded to the Galaxy platform and confirmed as having all adapters and low quality sequences removed by using FastQC (Afgan et al., 2018; Blankenberg et al., 2010). Before alignment, because sequences were developed using Illumina HiSeq2000, the RNAseq sequences were first converted to fastqsanger format using FASTQ Groomer (Blankenberg et al., 2010).

WL711 and TA5602 RNA reads were aligned against wheat chromosome 5D sequences downloaded from EnsemblPlants using HISAT2 (Hunt et al., 2018; Kim et al., 2015). Wheat chromosome 5D was used because *Ae. geniculata* chromosome 5M^g is highly similar to wheat chromosome 5D, with no major chromosomal rearrangements between the two chromosomes (Tiwari et al., 2015). Transcripts and

read counts were then reconstructed, normalized, and estimated by using StringTie both with and without the wheat 5D GFF3 file, also downloaded from EnsemblPlants (Kovaka et al., 2019; Pertea et al., 2015).

Gene count files were then used for differential expression analysis using both DESeq2 and edgeR (R. Liu et al., 2015; Love et al., 2014; Robinson et al., 2010). The criteria for differentially expressed genes were changes in $\log_2(\text{FC})$ and $P < 0.01$. Genes that were differentially expressed by both DESeq2 and edgeR were considered of particular interest.

***De novo* Assembly of TA5602 and WL711 Transcripts**

Trinity was used for *de novo* transcriptome assembly for both TA5602 and WL711 reads at 12 hours (Haas et al., 2013). The 12-hour time point was chosen because of Yadav et al., 2016 showing this is when differential gene expression would be the greatest. Trinity was run locally through a Docker container, to help ensure appropriate dependency configuration and Trinity execution (Merkel, 2014).

A TA10437-derived 5M^g scaffold was previously developed with paired end reads and 50 x coverage, and it was shown that 5M^g has high synteny and collinearity with wheat 5D and no major chromosomal rearrangements existed between 5D and 5M^g (Tiwari et al., 2015). As such, while it is certainly possible that particular orthologous may not exist between 5D and 5M^g, especially given the Chinese Spring reference is highly variable even against other modern wheat varieties, the general ordering of a 5M^g physical map can be inferred from 5D (Montenegro et al., 2017). High confidence 5D gene sequences were taken from 5D between 3.9 and 5 Mb and aligned to 5M^g

scaffold contigs via GMAP (Consortium (IWGSC) et al., 2018; Wu & Watanabe, 2005). The 5D-gene aligned 5M^s contigs were then ordered according to 5D gene order to create the physical map.

The *de novo* Trinity assemblies for 12-hour TA5602 and WL711 transcriptomes along with the 5M^s physical map before being aligned using HISAT2. Transcripts were then constructed based on the HISAT2 alignment using StringTie. Transcripts found through this process were examined by BLAST on IWGSC against the wheat reference sequence HighConfidenceGenesV1.1.

Results

High-Resolution Genetic Mapping and Identification of the *Lr57* Physical Region

Addition of a 5.1 Mb marker failed to reduce the genetic distance (2 cM) between the flanking markers of the *Lr57*. Using comparative genomics datasets such as GenomeZipper-based alignment across the sequenced cereal genomes a virtual gene map of chromosome 5M^s was generated earlier. Our group's work on chromosome 5M^s clearly established very high collinearity between 5M^s and 5D chromosomes, that means our 5M^s contigs can be aligned to the reference 5D of wheat. Additionally, it also suggested by using flanking markers of the *Lr57* region we can perform chromosomal landing of reference 5D and comparatively the 5M^s assembly. We also amplified all the genes in the comparative *Lr57* region and ordered them using our 5M^s genetic map. Our results again did not find any gene deletions or change in the order of any candidate genes in that interval. Since chromosomes 5M^s and 5D show such a close relationship, chromosome 5D-specific high-density SNP, POPSEQ and 5D physical map data (Wang et al. 2014, Chapman et al., 2015; Luo et al., 2013) were used

to anchor 5M^g genes and repeat-free NGS contigs (Tiwari et al, 2015). Anchoring and ordering of 5M^g contigs, gap filling using unmapped 5M^g reads, and sanger sequencing of candidate genes from the *Lr57* region enabled us to create a physical map of the *Lr57* region. Comparison of this physical region with 5D of wheat showed several SNPs and indels but gene order was collinear, and we did not observe any deletions in this physical region. The physical map of the *Lr57* region spanned a 1.1 Mb interval (3.97 to 5.00 Mb). Annotation of the 5M^g region and comparative analysis of the 5D reference map identified 7 candidate genes in the *Lr57* physical interval. To narrow down the number of candidate genes for final validation studies, we used gene ontology studies as well as transcriptome data to rule out unrelated candidate genes.


5D guided alignment of RNAseq reads

Our candidate region as delimited by the markers was a region spanning under 2 Mb, between what is assumed to be 3.9 and 5.5 Mb on 5M^g. Notably, only one NLR gene showed differential expression in this location, the wheat gene TraesCS5D02G005300, at 4 Mb. TraesCS5D02G005300 also showed the greatest gene expression change as reported by both edgeR and DESeq2 at all time points. These genes showed differential expression in the sense that $\log_2(\text{FC})$ was significantly large for WL711 at all time points (Table 4.1). This was because while reads were successfully aligning from WL711, no significant number of reads were aligning from TA5602 in our candidate region. Given WL711 is rust susceptible and we similarly saw no other notable TA5602 alignments in the 3.9 to 5.5 Mb region against wheat 5D,

we hypothesized that maybe 5M^g and 5D were too divergent at this location for mapping software to properly align RNA reads against 5D. There was also the possibility that 5D and 5M^g simply would not share an R-gene ortholog.

Table 4.1: DESeq2 Log2(FC) values after aligning RNA reads to the wheat 5D genome. Only the wheat gene TraesCS5D02G005300 consistently showed high differential gene expression between TA5602 and WL711 between 3.9 and 5.5 Mb, because only WL711 reads were successfully aligning. Some differences were seen also at TraesCS5D02G005600, but not to a significant degree

5D Gene Name	5D Location	Annotation	log2(FC)	log2(FC)	log2(FC)	log2(FC)	log2(FC)
TraesCS5D02G005300	4 mB	NLR	4.26	6.11	3.48	4.20	3.15
TraesCS5D02G005400	4.2 mB	NLR	NA	NA	NA	NA	NA
TraesCS5D02G005416	4.2 mB	pseudo NLR	NA	NA	NA	NA	NA
TraesCS5D02G005500	4.3 mB	Kinase	NA	NA	NA	NA	NA
TraesCS5D02G005600	4.3 mB	NLR	-0.14	-0.03	-0.08	-0.18	-0.14
p-adj < .01			0H	12H	24H	48H	72H



The 5M^g Physical Map and De Novo Assembly

A 5M^g physical map was created from a 5M^g scaffold with 50 x read coverage from Tiwari et al., 2015. This 5M^g map is 41.6 Kb long is equivalent to the 3.9 to 5 Mb 5D physical map. In this region, six high confidence 5D NLR genes and one 5D protein kinase could be aligned in close proximity (figure 4.1).

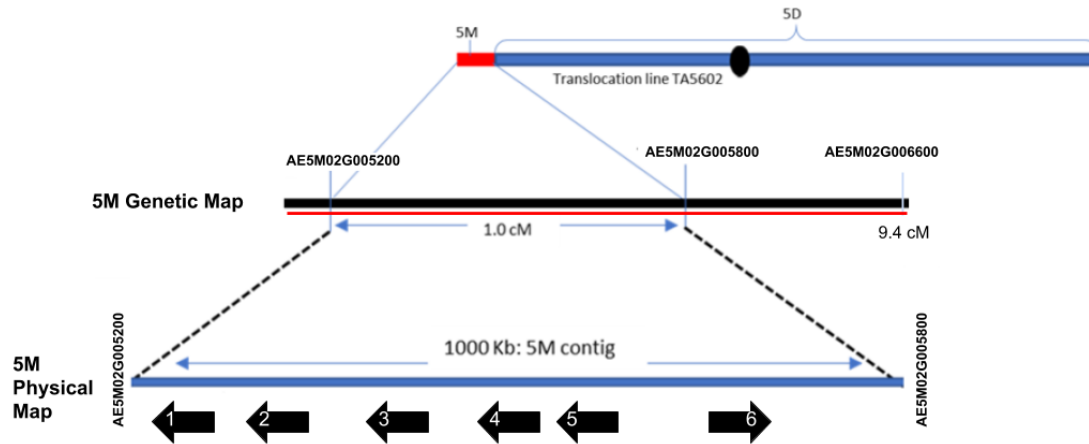


Figure 4.1: A representation of the mapped TA5602 introgression as it is represented by our genetic map and assembled physical map. Within the *Lr57* candidate gene region, there are seven 5D collinear genes as represented by the arrows 1-7 and inferred by 5D reference annotation- 5 NLR genes and 1 protein kinase (arrow 6). Gene number 1 is outside the *Lr57* region of interest.

To minimize potential bias of the 5D genome, Trinity assemblies were mapped against only the 5M^s physical map. By doing this, zero of the 92 thousand WL711 transcripts aligned against 5M^s. However, four transcripts of TA5602 aligned 5M^s. Two of the transcripts laid outside the *Lr57* mapping region. The other two transcripts aligned with overlap into what could form a single 852 Bp transcript. BLAST of this transcript against wheat 5D, 5B, and 5A showed alignment against the wheat NLR gene TraesCS5D02G005600 with a high identity and BLAST score (627/670 and 1012, respectively, with 12 gaps). Aligning 5D NLR genes against the 5M^s map, there is overlap of candidate NLR genes.

Attempting to align the *de novo* assemblies against the entire wheat 5D genome again showed only WL711 reads mapping within our *Lr57* region and only at TraesCS5D02G005300. Although leaf rust datasets are not available on the wheat expression browser, stripe rust study and PAMP studies are, and these also showed no

expression for any other gene than TraesCS5D02G005300 in response to being challenged with stripe rust or PAMPs.

Direct comparisons of 5D and 5M^s gene expression is somewhat complicated, given they are ultimately different and benefit the most from different types of alignment. At best we can compare the expression of the genes given their most favorable alignment conditions (Table 4.2). This does show that TA5602 generates uniquely mapping transcripts to TraesCS5D02G005600, and nothing else in the *Lr57* region from 3.9 to 5 Mb. Reads from 12-hour WL711, no matter if the reads are aligned to 5D or the 5M physical map, show no alignment to TraesCS5D02G005600.

For a more informative approach regarding TraesCS5D02G005600 expression in TA5602, we can also compare the expression of TA5602 after a mock treatment to the 12-hour treatment. This gives the most accurate view of TraesCS5D02G005600 potential as a candidate gene. To do this, Illumina adapter sequences were removed from the mock TA5602 RNA reads using Cutadapter and a Trinity de novo assembly was created as described above.

Comparing the mock TA5602 de novo assembly versus the 12 hour de novo assembly aligned against the 5M^s physical map with the same procedure as described above, we found that TraesCS5D02G005600 did not appear to be expressed in TA5602 after a mock treatment. If this is accurate, it shows TraesCS5D02G005600 has a $\log_2(\text{TPM}) = 18.9$ -fold increased expression change 12 hours after inoculation with leaf rust and this is unique compared to WL711.

Table 4.2: Comparing favorable mapping conditions for WL711 and TA5602. WL711 reads were aligned against the 5D reference and no expression was seen in TraesCS5D02G005600. WL711 reads were also not seen when aligning against the 5M^g physical map. Aligning TA5602 reads against the 5M^g physical map shows expression in a transcript with a high-percentage identity to TraesCS5D02G005600. This expression is absent in the mock treated TA5602.

5D location	5D Gene Name	Annotation	Log2(TPM)	Log2(TPM)
			WL711- 5D	TA5602- 5M
4 Mb	TraesCS5D02G005300	NB-ARC	4.9	0
4.2 Mb	TraesCS5D02G005400	NLR	0	0
4.2 Mb	TraesCS5D02G005416	pseudo- NLR	0	0
4.3 Mb	TraesCS5D02G005500	Protein kinase domain	0	0
4.3 Mb	TraesCS5D02G005600	NLR	0	18.9
5.1 Mb	TraesCS5D02G005700	PDZ domain	0	0

As of now this leaves us with one candidate gene between 3.9 and 5 Mb- TraesCS5D02G005600, an NLR gene that we now refer to as AE5M02G005600, given its presence in 5M^g. This is a reasonable candidate, both because of this transcription data, but also because it matches our earlier phenotyping and marker data well.

Discussion

From our genetic mapping, we had narrowed *Lr57* to a 2 cM region. From our phenotyping studies, it did appear that *Lr57* was likely to be an NLR-type gene. And in this region, there were 4 NLR genes and a protein kinase clustered together that we

had previously designed markers for that were polymorphic between resistant and susceptible *Ae. geniculata*. These markers also appeared to be linked to *Lr57*, but it was difficult to differentiate which of the genes should be our main candidate. To help solve this, we ultimately created a physical map of 5M^g by using our 5M^g scaffold contigs oriented and anchored against wheat 5D high confidence genes. This then allowed us to use previously generated RNAseq to identify a candidate gene for *Lr57*.

We wanted to revisit the RNAseq material generated in a study by our collaborators (Yadav et al., 2016), where WL711 versus TA5602 leaf rust response was analyzed before the release of the wheat 5D reference sequence. We were interested in how the 5D reference could extend our understanding of potential candidate genes in the mapped *Lr57* region. The original study by Yadav et al., (2016) implied NLR genes would likely be some of our candidates, and in our originally mapped small *Lr57* interval, we did see the typical phenomena of NLR genes clustering together- which can make choosing the correct NLR gene of interest more challenging (Michelson & Meyers, 1998). We would naturally expect that TA5602 would show increased expression of a candidate gene over WL711 after being challenged with a leaf rust pathogen. The difficulty would be in properly assigning the transcripts to genes. Reliable physical maps can be difficult to create for repetitive and large genomes, like those of the Triticeae and its relatives. We had access both to a scaffold of the 5M^g chromosome and now the high-quality wheat reference sequence of 5D as well. This gave us options for how we should try to align the RNAseq reads of TA5602. Either we could rely on the more complete 5D sequence with the hope that a wheat ortholog would be similar enough to our short RNA reads to allow proper

alignment, or we would need to first revisit the 5M^g scaffold to create a physical map in our region of interest and then align RNA reads. The latter option of creating a physical map had a clear benefit and necessity, because despite 5M^g and 5D being highly similar, it would be hard to confirm if we had missed 5M^g-unique genes that are absent in 5D if we only mapped RNA reads to 5D. Furthermore, TA5602 reads clearly would align more specifically to a 5M^g reference. But first attempting to map TA5602 RNA reads to the 5D reference had the advantage of 5D simply being readily available, complete, and already confirmed as being high quality. While perhaps mapping 5M^g to 5D would give less precise quantification of expression, there was still the possibility that this heuristic RNAseq approach would give us a workable candidate that could then be further verified by direct measurements, such as qPCR.

We found the method of aligning TA5602 RNA reads against 5D to be unreliable for mapping 5M^g transcripts, likely due to low shared sequence identity of short TA5602 5M^g RNA sequence reads against 5D. Or in a more difficult scenario, there was also the possibility the *Lr57* gene was only present uniquely in 5M^g, and that would prevent TA5602 RNA reads from aligning to our candidate region. As is, aligning the RNA reads against 5D showed only TraesCS5D02G005300, located approximately at 4 Mb, consistently had the highest differential gene expression as reported by both DESeq2 and EdgeR. But this expression was purely due to WL711 reads successfully aligning to this gene while no 5M^g could not be aligned, making the finding less informative and not explanatory of the *Lr57* phenotype. This showed

us that over-reliance on 5D homology to 5M^g would not be a workable option and aligning directly to 5D sequence was biasing our results too much.

We decided to create a *de novo* Trinity-based assembly for WL711 and TA5602, to remove 5D influence on our TA5602 RNA reads as much as possible. We also needed a 5M^g physical map. A 5M^g physical map would help find any genes that may be present in 5M^g but absent in WL711. However, we did not see it as necessary or practical to completely abandon the wealth of information the 5D reference provides- 5D and 5M^g still *are* highly similar, especially from the broader perspective of chromosome and gene arrangement, even if there's specific sequence differences in more localized regions. Because of this, we decided we would base 5M^g contig assembly into a physical map off of wheat 5D, due to the high collinearity between the genomes. High confidence 5D genes that have been shown to be within our mapped region of interest served as the basis for aligning 5M^g scaffold contigs via GMAP.

With the 5M^g physical map, aligning the *de novo* Trinity assemblies of WL711 and TA5602 with HISAT2 revealed only TA5602 transcripts that were aligning to the 5M^g physical map. Of these transcripts, only two overlapping transcripts were within the candidate region, and these two transcripts together formed one 800 Bp sequence. BLAST of this sequence against wheat showed high similarity to the wheat NLR gene, TraesCS5D02G005600 (4.3 Mb on 5D). But here is where comparing wheat versus *Ae. geniculata* RNA reads may come into question. 5M^g alignment clearly benefits accurately mapping 5M^g RNA reads, while 5D alignment clearly benefits WL711

reads. In our case, we are most interested first in understanding the function of our system of interest, 5M^g, so we necessarily prefer the method that accurately maps 5M^g RNA reads. Second, regardless if we attempt to map our 12-hour WL711 reads to 5D or 5M^g, it did not appear that WL711 significantly mapped to TraesCS5D02G005600, if at all. Even when comparing 5M^g aligned to 5D versus WL711 aligned to 5D, DESeq2 reported a small increased expression change in 5M^g, just not to any significant level. In either mapping scenario, with 5D or 5M^g, TraesCS5D02G005600 did not seem notably expressed from WL711. And none of the other genes between 3.9 and 5 Mb appeared to be notably expressed in either WL711 or TA5602. So instead our question became about how our candidate gene TraesCS5D02G005600, or AE5M02G005600 from its presence in 5M^g, might change in TA5602 from before versus after rust inoculation. This would give us a more accurate view of induced candidate genes. To do this, another de novo assembly of a mock TA5602 inoculation was created and aligned to the 5M^g physical map. The assembly of the mock treated TA5602 behaved similarly to WL711, and no transcripts aligned to our candidate gene AE5M2G005600 at a significant level. With this information, it appeared that 12 hours after inoculation with leaf rust, AE5M02G005600 in 5M^g has a Log₂(TPM) = 18.9-fold increased expression change.

This differential expression along with the hypersensitive rust responses that have been observed in TA5602 and TA5601 AE5M02G005600 is a good candidate for *Lr57*. Further verification methods of this gene will be done in future studies, including viral induced genome silencing (VIGS) and targeting induced local lesions IN genomes (TILLING) along with qPCR, overexpression, and CRISPR

transformation. As *Lr57* resistance has also been shown to persist at the adult plant stage, giving high levels of resistance, this is an exciting finding and shows the power of our unique mapping population that was used to originally map the *Lr57* candidate region. VIGS and qPCR validation of AE5M02G005600 are currently underway. We have also begun development of a TA5602 TILLING population.

From this thesis' work, we were able to show how a unique mapping population derived from an introgression line crossed with a disomic addition line allowed for efficient mapping of *Lr57* from wheat's tertiary gene pool. We were able to characterize the TA5602 5M^g introgression size and then further narrow down our region of interest due to the high mapping resolution of our population. Because of the wild type *Ph1* locus, recombination could be ensured to occur only within the small TA5602 introgression resistant for leaf rust and the homologous region of rust susceptible 5M^g chromatin in TA7659. The same techniques could be applied to any number of other genes from wheats tertiary or secondary gene pools, given the relative ease and speed of creating introgression lines. Accurate mapping and cloning of wild genes will hopefully be of great use in future breeding programs, both because of the huge degree of useful variation in wheat wild relatives, but also because accurate mapping will ensure reduced linkage drag that is typically associated with traits from wild relatives.

List of Primer Sequences used in the work

Position on 5D	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')
5Mg_60Mb	GTTCTTCTCATTGCCGTTTTCCT	ACTTGGTACCTTCATTTCATTGGC
5Mg_50Mb	AGAGGAAGAGGTTGAGGAGGGT	TATGGAACACGTTGGGGG
5Mg_45Mb	GATCGTACAAACAGGAGACGAG	TGACTTCTCTCGAGTCTGCAC
5Mg_35Mb	TGGCTTGACTTTTGACCCTAC	AAGGAATCGTTGACTCCTCTCTG
5Mg_17Mb	CCAGGTTTCTTCTTCTTCTTCT	GCGGCCTCTTCCAACATT
5Mg_9.9Mb	GCTGGAGATGCCCTTAGTACTCT	GCGAAATGATGGAGTCCTTG
5Mg_9.7Mb	TGACTCCAATAGGTTAGCCGAG	ATGGCTGCATGTGCGTGTA
5Mg_9.548Mb	AAATTGGGGGATGGGATAGG	GATTTTCATCCCCATAAACGCTAC
5Mg_8.9Mb	ATAGTCCCGTAGCATTTCCG	CGTCGATCGATGATAACCA
5Mg_8Mb	GTAAGTTATTGTGAGCGTCTGCG	AGAGAGCGTGTGATGAGTAGGAG
5Mg_7.8Mb	AACGCATGTATGGGTCACGA	GGTGGAAAGGTCAGTATGGCT
5Mg_7.4Mb	CGTTTACGAATTTTCATGACCTCT	TATCTGCTCAGGTGATGTTCTTG
5Mg_6Mb	CATTCACATTTGCAACGTGTACC	GCCACCAGTATACCATCTACATT
5Mg_5.6Mb	TGTTGAGCTTCTTTCTTTGTTTG	ATGTGCGAGAAGCATGTGGC
5Mg_5.5Mb	GCATTCTCCAATCAGTAGGCAAC	TGCCTGGCCAATGCATAA
5Mg_4.3Mb	GTCTCATTGAAGGCTCTCCTAAC	CATGACTTGTCTGCAATGAGG
5Mg_4.2Mb	GGCAGTTGGAAGGGCAAT	AAAATGTGAAGTCTGCAGAAGAG
5Mg_4Mb	CCGCAATGTTTCATTGGCTCC	ACTCGTCAACACAACAGTGCT
5Mg_3.9Mb	AATAAACAAGTGGGACCACAGA	AAGCAATCTTTGGTGGAACCTCAA
5Mg_3.6Mb	AAACTGAGCGCCATTGC	TCAATAATCCCAACCTGCAC
5Mg_1.79Mb	GGGTCAATCGACGGAATAAAA	AGGAGAAGAGAAGCTGCCG
5Mg_1.5Mb	AGTTAGGCTTTCAAAATTAGGGA	ATATATATACAACCGCGGCCTGA

Bibliography

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- Aghaee-Sarbarzeh, M., Ferrahi, M., Singh, S., Singh, H., Friebe, B., Gill, B. S., & Dhaliwal, H. S. (2002). *PhI -induced transfer of leaf and stripe rust-resistance genes from Aegilops triuncialis and Ae. Geniculata to bread wheat*. 6.
- Ahn, S., Anderson, J. A., Sorrells, M. E., & Tanksley, S. 1993. (1993). Homoeologous relationships of rice, wheat and maize chromosomes. *Molecular and General Genetics MGG*, 241(5–6), 483–490.
- Aime, M. C., McTaggart, A. R., Mondo, S. J., & Duplessis, S. (2017). *Chapter Seven-Phylogenetics and Phylogenomics of Rust Fungi*. In: *Fungal Phylogenetics and Phylogenomics*. Lawrence Berkeley National Laboratory-National Energy Research Scientific
- Ali, M. (2010). *Years of Drought: A Report on the Effects of Drought on the Syrian Peninsula*. 14.
- Ali, S., Rodriguez-Algaba, J., Thach, T., Sørensen, C. K., Hansen, J. G., Lassen, P., Nazari, K., Hodson, D. P., Justesen, A. F., & Hovmøller, M. S. (2017). Yellow rust epidemics worldwide were caused by pathogen races from divergent genetic lineages. *Frontiers in Plant Science*, 8, 1057.
- Al-Kaff, N., Knight, E., Bertin, I., Foote, T., Hart, N., Griffiths, S., & Moore, G. (2008). Detailed dissection of the chromosomal region containing the Ph1 locus in wheat *Triticum aestivum*: With deletion mutants and expression profiling. *Annals of Botany*, 101(6), 863–872.
- Amil, R. E., Ali, S., Bahri, B., Leconte, M., Vallavieille-Pope, C. de, & Nazari, K. (2020). Pathotype diversification in the invasive PstS2 clonal lineage of *Puccinia striiformis* f. Sp. *Tritici* causing yellow rust on durum and bread wheat in Lebanon and Syria in 2010–2011. *Plant Pathology*, 69(4), 618–630. <https://doi.org/10.1111/ppa.13164>
- Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11(10), R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
- Arrigo, N., Felber, F., Parisod, C., Buerki, S., Alvarez, N., David, J., & Guadagnuolo, R. (2010). Origin and expansion of the allotetraploid *Aegilops geniculata*, a wild relative of wheat. *New Phytologist*, 187(4), 1170–1180. <https://doi.org/10.1111/j.1469-8137.2010.03328.x>
- Bakkeren, G., & Szabo, L. J. (2019). Progress on Molecular Genetics and Manipulation of Rust Fungi. *Phytopathology®*, 110(3), 532–543. <https://doi.org/10.1094/PHYTO-07-19-0228-IA>

- Bansal, M., Kaur, S., Dhaliwal, H. S., Bains, N. S., Bariana, H. S., Chhuneja, P., & Bansal, U. K. (2017). Mapping of *Aegilops umbellulata*-derived leaf rust and stripe rust resistance loci in wheat. *Plant Pathology*, 66(1), 38–44.
- Bansal, Mitaly, Adamski, N. M., Toor, P. I., Kaur, S., Molnár, I., Holušová, K., Vrána, J., Doležel, J., Valárik, M., Uauy, C., & Chhuneja, P. (2020). *Aegilops umbellulata* introgression carrying leaf rust and stripe rust resistance genes Lr76 and Yr70 located to 9.47-Mb region on 5DS telomeric end through a combination of chromosome sorting and sequencing. *Theoretical and Applied Genetics*, 133(3), 903–915. <https://doi.org/10.1007/s00122-019-03514-x>
- Barker, G., & Goucher, C. (2015). *The Cambridge World History: Volume 2, A World with Agriculture, 12,000 BCE–500 CE*. Cambridge University Press.
- Beddow, J. M., Pardey, P. G., Chai, Y., Hurley, T. M., Kriticos, D. J., Braun, H.-J., Park, R. F., Cuddy, W. S., & Yonow, T. (2015). Research investment implications of shifts in the global geography of wheat stripe rust. *Nature Plants*, 1(10), 1–5. <https://doi.org/10.1038/nplants.2015.132>
- Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., & Nekrutenko, A. (2010). Manipulation of FASTQ data with Galaxy. *Bioinformatics*, 26(14), 1783–1785. <https://doi.org/10.1093/bioinformatics/btq281>
- Bs, G., Hc, S., Wj, R., Le, B., Jh, H., Tl, H., Jg, M., & Jg, W. (1985). Evaluation of *Aegilops* species for resistance to wheat powdery mildew, wheat leaf rust, hessian fly, and greenbug. *Plant Disease*, 69(4), 314–316.
- Chen, P. D., Tsujimoto, H., & Gill, B. S. (1994). *Transfer of Ph genes promoting homoeologous pairing from Triticum speltoides to common wheat*. 5.
- Chhuneja, P., Kaur, S., Goel, R. K., Aghaee-Sarbarzeh, M., Prashar, M., & Dhaliwal, H. S. (2008). Transfer of leaf rust and stripe rust resistance from *Aegilops umbellulata* Zhuk. To bread wheat (*Triticum aestivum* L.). *Genetic Resources and Crop Evolution*, 55(6), 849–859.
- Compeau, P. E. C., Pevzner, P. A., & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology*, 29(11), 987–991. <https://doi.org/10.1038/nbt.2023>
- Consortium (IWGSC), T. I. W. G. S., Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C. J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Sharpe, A. G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A.-A., ... Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403). <https://doi.org/10.1126/science.aar7191>
- Devos, K. M., & Gale, M. D. (2000). Genome relationships: The grass model in current research. *The Plant Cell*, 12(5), 637–646.
- Dhaliwal, H. S., Singh, H., Gill, K. S., & Randhawa, H. S. (1993). Evaluation and cataloguing of wheat germplasm for disease resistance and quality. *Biodiversity and Wheat Improvement*, 123–140.
- Dondlinger, P. T. (1919). *The Book of Wheat: An Economic History and Practical Manual of the Wheat Industry*. Orange Judd Company.
- Ellis, J. G., Lagudah, E. S., Spielmeier, W., & Dodds, P. N. (2014). The past, present and future of breeding rust resistant wheat. *Frontiers in Plant Science*, 5. <https://doi.org/10.3389/fpls.2014.00641>

- Enghiad, A., Ufer, D., Countryman, A. M., & Thilmany, D. D. (2017). An Overview of Global Wheat Market Fundamentals in an Era of Climate Concerns. *International Journal of Agronomy*, 2017, 1–15. <https://doi.org/10.1155/2017/3931897>
- FAOSTAT. (n.d.-a). Retrieved June 7, 2020, from <http://www.fao.org/faostat/en/#data/FBS>
- FAOSTAT. (n.d.-b). Retrieved June 7, 2020, from <http://www.fao.org/faostat/en/#data/QC>
- Faris, J. D. (2014). Wheat Domestication: Key to Agricultural Revolutions Past and Future. In R. Tuberosa, A. Graner, & E. Frison (Eds.), *Genomics of Plant Genetic Resources* (pp. 439–464). Springer Netherlands. https://doi.org/10.1007/978-94-007-7572-5_18
- Figueroa, M., Hammond-Kosack, K. E., & Solomon, P. S. (2018). A review of wheat diseases—A field perspective. *Molecular Plant Pathology*, 19(6), 1523–1536. <https://doi.org/10.1111/mpp.12618>
- Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, 9(1), 275–296.
- Forsström, P.-O., & Merker, A. (2001). Sources of wheat powdery mildew resistance from wheat-rye and wheat-Leymus hybrids. *Hereditas*, 134(2), 115–119.
- Friebe, B., Jiang, J., Raupp, W. J., McIntosh, R. A., & Gill, B. S. (1996). Characterization of wheat-alien translocations conferring resistance to diseases and pests: Current status. *Euphytica*, 91(1), 59–87. <https://doi.org/10.1007/BF00035277>
- Friebe, B. R., Tuleen, N. A., & Gill, B. S. (1999). *Development and identification of a complete set of Triticum aestivum – Aegilops geniculata chromosome addition lines*. 42, 7.
- Gill, K. S., Gill, B. S., & Endo, T. R. (1993). A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. *Chromosoma*, 102(6), 374–381.
- Gill, Khem Singh, Dhaliwal, H. S., & Singh, H. (1995). *Cataloguing and pre-breeding of wheat genetic resources*.
- Gill, Kulvinder S., Gill, B. S., Endo, T. R., & Boyko, E. V. (1996). Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. *Genetics*, 143(2), 1001–1012.
- Gill, B. S., Friebe, B., Raupp, W. J., Wilson, D. L., Cox, T. S., Sears, R. G., Brown-Guedira, G. L., & Fritz, A. K. (2006). Wheat genetics resource center: The first 25 years. *Advances in Agronomy*, 89, 73–136.
- Griffiths, S., Sharp, R., Foote, T. N., Bertin, I., Wanous, M., Reader, S., Colas, I., & Moore, G. (2006). Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature*, 439(7077), 749–752.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>

- Harjit, S., & Dhaliwal, H. S. (2000). Intraspecific genetic diversity for resistance to wheat rusts in wild *Triticum* and *Aegilops* species. *Wheat Information Service*, 90, 21–30.
- Hedden, P. (2003). The genes of the Green Revolution. *Trends in Genetics*, 19(1), 5–9. [https://doi.org/10.1016/S0168-9525\(02\)00009-4](https://doi.org/10.1016/S0168-9525(02)00009-4)
- Hovmøller, M. S., & Justesen, A. F. (2007). Rates of evolution of avirulence phenotypes and DNA markers in a northwest European population of *Puccinia striiformis* f. Sp. *Tritici*. *Molecular Ecology*, 16(21), 4637–4647.
- Huang, L., Brooks, S. A., Li, W., Fellers, J. P., Trick, H. N., & Gill, B. S. (2003). Map-based cloning of leaf rust resistance gene Lr21 from the large and polyploid genome of bread wheat. *Genetics*, 164(2), 655–664.
- Huerta-Espino, J., Singh, R. P., Germán, S., McCallum, B. D., Park, R. F., Chen, W. Q., Bhardwaj, S. C., & Goyeau, H. (2011). Global status of wheat leaf rust caused by *Puccinia triticina*. *Euphytica*, 179(1), 143–160. <https://doi.org/10.1007/s10681-011-0361-x>
- Hunt, S. E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., Armean, I. M., Trevanion, S. J., Flicek, P., & Cunningham, F. (2018). Ensembl variation resources. *Database*, 2018. <https://doi.org/10.1093/database/bay119>
- Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W., & Mortensen, D. A. (2017). Agriculture in 2050: Recalibrating targets for sustainable intensification. *Bioscience*, 67(4), 386–391.
- Hurni, S., Brunner, S., Buchmann, G., Herren, G., Jordan, T., Krukowski, P., Wicker, T., Yahiaoui, N., Mago, R., & Keller, B. (2013). Rye P m8 and wheat P m3 are orthologous genes and show evolutionary conservation of resistance function against powdery mildew. *The Plant Journal*, 76(6), 957–969.
- Jiang, J., Friebe, B., & Gill, B. S. (1993). Recent advances in alien gene transfer in wheat. *Euphytica*, 73(3), 199–212.
- Johal, G. S., & Briggs, S. P. (1992). Reductase activity encoded by the HM1 disease resistance gene in maize. *Science*, 258(5084), 985–987.
- Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323–329. <https://doi.org/10.1038/nature05286>
- Jordan, T., Seeholzer, S., Schwizer, S., Töller, A., Somssich, I. E., & Keller, B. (2011). The wheat Mla homologue TmMla1 exhibits an evolutionarily conserved function against powdery mildew in both wheat and barley. *The Plant Journal*, 65(4), 610–621.
- Kalia, B. (2015). *Mining the Aegilops tauschii gene pool: Evaluation, introgression and molecular characterization of adult plant resistance to leaf rust and seedling resistance to tan spot in synthetic hexaploid wheat*. Kansas State University.
- Keller, B., & Feuillet, C. (2000). Colinearity and gene density in grass genomes. *Trends in Plant Science*, 5(6), 246–251.
- Kelley, C. P., Mohtadi, S., Cane, M. A., Seager, R., & Kushnir, Y. (2015). Climate change in the Fertile Crescent and implications of the recent Syrian drought. *Proceedings of the National Academy of Sciences*, 112(11), 3241–3246. <https://doi.org/10.1073/pnas.1421533112>

- Kilian, B., Mammen, K., Millet, E., Sharma, R., Graner, A., Salamini, F., Hammer, K., & Özkan, H. (2011). *Aegilops*. In C. Kole (Ed.), *Wild Crop Relatives: Genomic and Breeding Resources* (pp. 1–76). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14228-4_1
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kishii, M. (2019). An Update of Recent Use of *Aegilops* Species in Wheat Breeding. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.00585>
- Kolmer, J. (2013). Leaf rust of wheat: Pathogen biology, variation and host resistance. *Forests*, 4(1), 70–84.
- Koo, D.-H., Liu, W., Friebe, B., & Gill, B. S. (2017). Homoeologous recombination in the presence of Ph1 gene in wheat. *Chromosoma*, 126(4), 531–540. <https://doi.org/10.1007/s00412-016-0622-5>
- Kourelis, J., & Hoorn, R. A. L. van der. (2018). Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. *The Plant Cell*, 30(2), 285–299. <https://doi.org/10.1105/tpc.17.00579>
- Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., & Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*, 20(1), 278. <https://doi.org/10.1186/s13059-019-1910-1>
- Kumar, S. (2020). *Fine mapping and expression analysis of stripe rust resistance genes derived from Aegilops geniculata*. Punjab Agricultural University, Ludhiana.
- Kuraparthi, V. (2007). *Genomic targeting and mapping of agronomically important genes in wheat*. <https://krex.k-state.edu/dspace/handle/2097/311>
- Kuraparthi, V., Chhuneja, P., Dhaliwal, H. S., Kaur, S., Bowden, R. L., & Gill, B. S. (2007). Characterization and mapping of cryptic alien introgression from *Aegilops geniculata* with new leaf rust and stripe rust resistance genes Lr57 and Yr40 in wheat. *Theoretical and Applied Genetics*, 114(8), 1379–1389. <https://doi.org/10.1007/s00122-007-0524-2>
- Kuraparthi, V., Sood, S., & Gill, B. S. (2009). Molecular genetic description of the cryptic wheat-*Aegilops geniculata* introgression carrying rust resistance genes Lr57 and Yr40 using wheat ESTs and synteny with rice. *Genome*, 52(12), 1025–1036. <https://doi.org/10.1139/G09-076>
- Ladizinsky, G. (1985). Founder effect in crop-plant evolution. *Economic Botany*, 39(2), 191–199. <https://doi.org/10.1007/BF02907844>
- Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., & Wang, J. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*, 20(2), 265–272. <https://doi.org/10.1101/gr.097261.109>
- Liu, R., Holik, A. Z., Su, S., Jansz, N., Chen, K., Leong, H. S., Blewitt, M. E., Asselin-Labat, M.-L., Smyth, G. K., & Ritchie, M. E. (2015). Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Research*, 43(15), e97–e97. <https://doi.org/10.1093/nar/gkv412>
- Liu, W., Rouse, M., Friebe, B., Jin, Y., Gill, B., & Pumphrey, M. O. (2011). Discovery

- and molecular mapping of a new gene conferring resistance to stem rust, Sr53, derived from *Aegilops geniculata* and characterization of spontaneous translocation stocks with reduced alien chromatin. *Chromosome Research*, 19(5), 669–682. <https://doi.org/10.1007/s10577-011-9226-3>
- Loureiro, I., Escorial, M. C., García-Baudín, J. M., & Chueca, M. C. (2006). Evidence of natural hybridization between *Aegilops geniculata* and wheat under field conditions in Central Spain. *Environmental Biosafety Research*, 5(2), 105–109. <https://doi.org/10.1051/ebr:2006020>
- Loureiro, I., Escorial, M. C., García-Baudin, J. M., & Chueca, M. C. (2007). Hybridization between wheat (*Triticum aestivum*) and the wild species *Aegilops geniculata* and *A. biuncialis* under experimental field conditions. *Agriculture, Ecosystems & Environment*, 120(2), 384–390. <https://doi.org/10.1016/j.agee.2006.10.015>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Lu, F., Romay, M. C., Glaubitz, J. C., Bradbury, P. J., Elshire, R. J., Wang, T., Li, Y., Li, Y., Semagn, K., Zhang, X., Hernandez, A. G., Mikel, M. A., Soifer, I., Barad, O., & Buckler, E. S. (2015). High-resolution genetic mapping of maize pan-genome sequence anchors. *Nature Communications*, 6(1), 6914. <https://doi.org/10.1038/ncomms7914>
- Lukaszewski, A. J., & Curtis, C. A. (1993). Physical distribution of recombination in B-genome chromosomes of tetraploid wheat. *Theoretical and Applied Genetics*, 86(1), 121–127.
- Lukaszewski, A. J. (1995). Physical distribution of translocation breakpoints in homoeologous recombinants induced by the absence of the Ph1 gene in wheat and triticale. *Theoretical and Applied Genetics*, 90(5), 714–719.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., ... Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1, 18. <https://doi.org/10.1186/2047-217X-1-18>
- Mago, R., Zhang, P., Vautrin, S., Šimková, H., Bansal, U., Luo, M.-C., Rouse, M., Karaoglu, H., Periyannan, S., & Kolmer, J. (2015). The wheat Sr50 gene reveals rich diversity at a cereal disease resistance locus. *Nature Plants*, 1(12), 1–3.
- Markell, S. G., & Milus, E. A. (2008). Emergence of a novel population of *Puccinia striiformis* f. Sp. *Tritici* in eastern United States. *Phytopathology*, 98(6), 632–639.
- Martín, A. C., Rey, M.-D., Shaw, P., & Moore, G. (2017). Dual effect of the wheat Ph1 locus on chromosome synapsis and crossover. *Chromosoma*, 126(6), 669–680. <https://doi.org/10.1007/s00412-017-0630-0>
- Martín, A. C., Shaw, P., Phillips, D., Reader, S., & Moore, G. (2014). Licensing MLH1 sites for crossover during meiosis. *Nature Communications*, 5(1), 4580. <https://doi.org/10.1038/ncomms5580>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10–12.

<https://doi.org/10.14806/ej.17.1.200>

- Merkel, D. (2014). Docker: Lightweight linux containers for consistent development and deployment. *Linux Journal*, 2014(239), 2.
- Michelmore, R. W., & Meyers, B. C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research*, 8(11), 1113–1130.
- Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C.-K. K., Visendi, P., Lai, K., Doležel, J., & Batley, J. (2017). The pangenome of hexaploid bread wheat. *The Plant Journal*, 90(5), 1007–1013.
- Neelam, K., Rawat, N., Tiwari, V. K., Kumar, S., Chhuneja, P., Singh, K., Randhawa, G. S., & Dhaliwal, H. S. (2011). Introgression of group 4 and 7 chromosomes of *Ae. Peregrina* in wheat enhances grain iron and zinc density. *Molecular Breeding*, 28(4), 623–634.
- Ohta, S. (2017). Diverse morphological and cytogenetic variation and differentiation of the two subspecies in *Aegilops geniculata* Roth, a wild relative of wheat. *Genetic Resources and Crop Evolution*, 64(8), 2009–2020. <https://doi.org/10.1007/s10722-017-0492-6>
- OKAMOTO, M. (1957). Asynaptic effect of chromosome V. *Wheat Inf. Serv.*, 5, 6.
- Oliver, R. P. (2014). A reassessment of the risk of rust fungi developing resistance to fungicides. *Pest Management Science*, 70(11), 1641–1645. <https://doi.org/10.1002/ps.3767>
- Olivera, P. D., Rouse, M. N., & Jin, Y. (2018). Identification of New Sources of Resistance to Wheat Stem Rust in *Aegilops* spp. In the Tertiary Genepool of Wheat. *Frontiers in Plant Science*, 9. <https://doi.org/10.3389/fpls.2018.01719>
- Peng, F. Y., & Yang, R.-C. (2017). Prediction and analysis of three gene families related to leaf rust (*Puccinia triticina*) resistance in wheat (*Triticum aestivum* L.). *BMC Plant Biology*, 17(1), 108.
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290–295. <https://doi.org/10.1038/nbt.3122>
- Pingali, P. L. (2012). Green Revolution: Impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences*, 109(31), 12302–12308. <https://doi.org/10.1073/pnas.0912953109>
- Ponce-Molina, L. J., Huerta-Espino, J., Singh, R. P., Basnet, B. R., Alvarado, G., Randhawa, M. S., Lan, C. X., Aguilar-Rincón, V. H., Lobato-Ortiz, R., & García-Zavala, J. J. (2018). Characterization of Leaf Rust and Stripe Rust Resistance in Spring Wheat ‘Chilero.’ *Plant Disease*, 102(2), 421–427. <https://doi.org/10.1094/PDIS-11-16-1545-RE>
- Prasad, P., Savadi, S., Bhardwaj, S. C., & Gupta, P. K. (2020). The progress of leaf rust research in wheat. *Fungal Biology*, 124(6), 537–550. <https://doi.org/10.1016/j.funbio.2020.02.013>
- PSD Online. (n.d.). Retrieved June 7, 2020, from <https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery>
- Qi, L., Friebe, B., Zhang, P., & Gill, B. S. (2007). Homoeologous recombination, chromosome engineering and crop improvement. *Chromosome Research*,

- 15(1), 3–19. <https://doi.org/10.1007/s10577-006-1108-8>
- Qi, L. L., Echalier, B., Chao, S., Lazo, G. R., Butler, G. E., Anderson, O. D., Akhunov, E. D., Dvořák, J., Linkiewicz, A. M., Ratnasiri, A., Dubcovsky, J., Bermudez-Kandianis, C. E., Greene, R. A., Kantety, R., Rota, C. M. L., Munkvold, J. D., Sorrells, S. F., Sorrells, M. E., Dilbirligi, M., ... Gill, B. S. (2004). A Chromosome Bin Map of 16,000 Expressed Sequence Tag Loci and Distribution of Genes Among the Three Genomes of Polyploid Wheat. *Genetics*, 168(2), 701–712. <https://doi.org/10.1534/genetics.104.034868>
- Rawat, N., Tiwari, V. K., Singh, N., Randhawa, G. S., Singh, K., Chhuneja, P., & Dhaliwal, H. S. (2008). Evaluation and utilization of Aegilops and wild Triticum species for enhancing iron and zinc content in wheat. *Genetic Resources and Crop Evolution*, 56(1), 53. <https://doi.org/10.1007/s10722-008-9344-8>
- Rey, M.-D., Calderón, M. C., & Prieto, P. (2015). The use of the ph1b mutant to induce recombination between the chromosomes of wheat and barley. *Frontiers in Plant Science*, 6, 160.
- Rey, M.-D., Martín, A. C., Higgins, J., Swarbreck, D., Uauy, C., Shaw, P., & Moore, G. (2017). Exploiting the ZIP4 homologue within the wheat Ph1 locus has identified two lines exhibiting homoeologous crossover in wheat-wild relative hybrids. *Molecular Breeding*, 37(8), 95. <https://doi.org/10.1007/s11032-017-0700-2>
- Riley, R., & Chapman, V. (1958). Genetic Control of the Cytologically Diploid Behaviour of Hexaploid Wheat. *Nature*, 182(4637), 713–715. <https://doi.org/10.1038/182713a0>
- Riley, R., Chapman, V., & Johnson, R. O. Y. (1968). Introduction of yellow rust resistance of Aegilops comosa into wheat by genetically induced homoeologous recombination. *Nature*, 217(5126), 383–384.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Roelfs, A. P. (1992). *Rust diseases of wheat: Concepts and methods of disease management*. Cimmyt.
- Sánchez-Morán, E., Benavente, E., & Orellana, J. (2001). Analysis of karyotypic stability of homoeologous-pairing (ph) mutants in allopolyploid wheats. *Chromosoma*, 110(5), 371–377. <https://doi.org/10.1007/s004120100156>
- Schatz, M. C., Delcher, A. L., & Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. *Genome Research*, 20(9), 1165–1173. <https://doi.org/10.1101/gr.101360.109>
- Schneider, A., Molnár, I., & Molnár-Láng, M. (2008). Utilisation of Aegilops (goatgrass) species to widen the genetic diversity of cultivated wheat. *Euphytica*, 163(1), 1–19. <https://doi.org/10.1007/s10681-007-9624-y>
- Schwessinger, B. (2017). Fundamental wheat stripe rust research in the 21st century. *New Phytologist*, 213(4), 1625–1631. <https://doi.org/10.1111/nph.14159>
- Schwessinger, B., Chen, Y.-J., Tien, R., Vogt, J. K., Sperschneider, J., Nagar, R., McMullan, M., Sicheritz-Ponten, T., Sørensen, C. K., & Hovmøller, M. S. (2020). Distinct life histories impact dikaryotic genome evolution in the rust

- fungus *Puccinia striiformis* causing stripe rust in wheat. *Genome Biology and Evolution*, 12(5), 597–617.
- Sears, E. R., & Okamoto, M. (1958). Intergenomic chrchanosome pairing in hexploid wheat. *Proc. 10th Intern. Congr Genet*, 2, 258–259.
- Steele, K. A., Humphreys, E., Wellings, C. R., & Dickinson, M. J. (2001). Support for a stepwise mutation model for pathogen evolution in Australasian *Puccinia striiformis* f. Sp. *Triticici* by use of molecular markers. *Plant Pathology*, 50(2), 174–180.
- Tarr, D. E. K., & Alexander, H. M. (2009). TIR-NBS-LRR genes are rare in monocots: Evidence from diverse monocot orders. *BMC Research Notes*, 2(1), 197. <https://doi.org/10.1186/1756-0500-2-197>
- Tavares, S., Ramos, A. P., Pires, A. S., Azinheira, H. G., Caldeirinha, P., Link, T., Abranches, R., Silva, M. do C., Voegelé, R. T., Loureiro, J., & Talhinhos, P. (2014). Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in Plant Science*, 5. <https://doi.org/10.3389/fpls.2014.00422>
- Tiwari, V. K., Wang, S., Danilova, T., Koo, D. H., Vrána, J., Kubaláková, M., Hribova, E., Rawat, N., Kalia, B., Singh, N., Friebe, B., Doležal, J., Akhunov, E., Poland, J., Sabir, J. S. M., & Gill, B. S. (2015). Exploring the tertiary gene pool of bread wheat: Sequence assembly and analysis of chromosome 5Mg of *Aegilops geniculata*. *The Plant Journal*, 84(4), 733–746. <https://doi.org/10.1111/tpj.13036>
- Tiwari, V. K., Wang, S., Sehgal, S., Vrána, J., Friebe, B., Kubaláková, M., Chhuneja, P., Doležal, J., Akhunov, E., Kalia, B., Sabir, J., & Gill, B. S. (2014). SNP Discovery for mapping alien introgressions in wheat. *BMC Genomics*, 15(1), 273. <https://doi.org/10.1186/1471-2164-15-273>
- Trigo, R. M., Gouveia, C. M., & Barriopedro, D. (2010). The intense 2007–2009 drought in the Fertile Crescent: Impacts and associated atmospheric circulation. *Agricultural and Forest Meteorology*, 150(9), 1245–1257. <https://doi.org/10.1016/j.agrformet.2010.05.006>
- United Nations, Department of Economic and Social Affairs, & Population Division. (2019). *World population prospects Highlights, 2019 revision Highlights, 2019 revision*.
- University, U. N., & Organization, W. H. (2004). *Human Energy Requirements: Report of a Joint FAO/WHO/UNU Expert Consultation: Rome, 17-24 October 2001* (Vol. 1). Food & Agriculture Org.
- USDA ERS - Wheat Data. (n.d.). Retrieved June 7, 2020, from <https://www.ers.usda.gov/data-products/wheat-data/>
- Valkoun, J., Hammer, K., Kučerová, D., & Bartoš, P. (1985). Disease resistance in the genus *Aegilops* L.—Stem rust, leaf rust, stripe rust, and powdery mildew. *Die Kulturpflanze*, 33(2), 133–153.
- Wang, M., & Chen, X. (2017). Stripe rust resistance. In *Stripe rust* (pp. 353–558). Springer.
- Wang, S., Yin, L., Tanaka, H., Tanaka, K., & Tsujimoto, H. (2011). Wheat-*Aegilops* chromosome addition lines showing high iron and zinc contents in grains. *Breeding Science*, 61(2), 189–195.
- Weiss, E., & Zohary, D. (2011). The Neolithic Southwest Asian Founder Crops: Their

- Biology and Archaeobotany. *Current Anthropology*, 52(S4), S237–S254. <https://doi.org/10.1086/658367>
- Wik, M., Pingali, P., & Brocai, S. (2008). *Global agricultural performance: Past trends and future prospects*.
- Wilson, S. (1873). II. Wheat and Rye Hybrids. *Transactions of the Botanical Society of Edinburgh*, 12(1–4), 286–288.
- Wu, T. D., & Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9), 1859–1875.
- Yadav, I. S., Sharma, A., Kaur, S., Nahar, N., Bhardwaj, S. C., Sharma, T. R., & Chhuneja, P. (2016). Comparative Temporal Transcriptome Profiling of Wheat near Isogenic Line Carrying Lr57 under Compatible and Incompatible Interactions. *Frontiers in Plant Science*, 7. <https://doi.org/10.3389/fpls.2016.01943>
- Zaharieva, M., Monneveux, P., Henry, M., Rivoal, R., Valkoun, J., & Nachit, M. M. (2001). Evaluation of a Collection of Wild Wheat Relative *Aegilops Geniculata* Roth and Identification of Potential Sources for Useful Traits. In Z. Bedö & L. Láng (Eds.), *Wheat in a Global Environment: Proceedings of the 6th International Wheat Conference, 5–9 June 2000, Budapest, Hungary* (pp. 739–746). Springer Netherlands. https://doi.org/10.1007/978-94-017-3674-9_99
- Zhao, C. Z., Li, Y. H., Dong, H. T., Geng, M. M., Liu, W. H., Li, F., Ni, Z. F., Wang, X. J., Xie, C. J., & Sun, Q. X. (2016). Molecular cloning, functional verification, and evolution of TmPm3, the powdery mildew resistance gene of *Triticum monococcum* L. *Genet Mol Res*, 15(2), gmr. 15028056.
- Zhao, G., Zou, C., Li, K., Wang, K., Li, T., Gao, L., Zhang, X., Wang, H., Yang, Z., Liu, X., Jiang, W., Mao, L., Kong, X., Jiao, Y., & Jia, J. (2017). The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nature Plants*, 3(12), 946–955. <https://doi.org/10.1038/s41477-017-0067-8>
- Zhao, J., Wang, M., Chen, X., & Kang, Z. (2016). Role of alternate hosts in epidemiology and pathogen variation of cereal rusts. *Annual Review of Phytopathology*, 54, 207–228.
- Zickler, D., & Kleckner, N. (1999). Meiotic Chromosomes: Integrating Structure and Function. *Annual Review of Genetics*, 33(1), 603–754. <https://doi.org/10.1146/annurev.genet.33.1.603>
- Zipfel, C. (2014). Plant pattern-recognition receptors. *Trends in Immunology*, 35(7), 345–351. <https://doi.org/10.1016/j.it.2014.05.004>